

Downlink Performance of Pilot-Reused HetNet with Large-Scale Antenna Arrays

Yongyu Dai, *Student Member, IEEE*, Xiaodai Dong, *Senior Member, IEEE*, and Hai Lin, *Senior Member, IEEE*

Abstract—Considering a heterogeneous network (HetNet) where both macro base station (BS) and small cell (SC) nodes are equipped with massive antennas, this paper studies the performance for multiple-input multiple-output (MIMO) downlinks when the macro and small cells share the same spectrum and hence interfere with each other. Suppose that the large-scale antenna arrays at both macro BS and SC nodes employ maximum-ratio transmission (MRT) or zero-forcing transmission (ZFT) precoding, and transmit data streams to the served users simultaneously. A new pilot reuse pattern among SCs is proposed for channel estimation. Taking into account imperfect channel state information (CSI), capacity lower bounds for MRT and ZFT are derived, respectively, in closed-form expressions involving only statistical CSI. Then asymptotic analyses for massive arrays are presented under specific power scaling laws. Subsequently, two user scheduling algorithms, greedy scheduling algorithm and asymptotical scheduling algorithm (ASA), are proposed based on derived capacity lower bounds and asymptotic analyses, respectively. ASA is demonstrated to be a near optimal in the asymptotic regime and has low complexity. Finally, the derived closed-form expressions are verified to be accurate predictors of the system performance by Monte-Carlo simulations. Numerical results demonstrate the effectiveness of asymptotic analysis and proposed user scheduling schemes.

Index Terms—Heterogeneous network, MIMO, large-scale antenna arrays, pilot reuse, user scheduling.

I. INTRODUCTION

As a viable and cost-effective way to increase network capacity, heterogeneous networks (HetNets) that embed a large number of low-power nodes, called small cells (SCs), into an existing macro network has emerged with the aim to offload traffic from the macro cell (MC) to small cells [2–6] in hot spots or to solve coverage holes in MC. Conventionally deploying more macro base stations (BSs) in already dense networks may be prohibitively expensive and result in severe inter-cell interference [5]. However, due to the large number of potentially interfering nodes in the network, mitigating both the inter-cell and intra-cell interference becomes a crucial issue facing HetNet. Interference control has been intensively studied and applied in HetNet [7–9], including the coordinated multi-point (CoMP) transmission [7]. Although the CoMP transmission was shown to provide high spectral efficiency [10] with the backhaul among the coordinated tiers enabling both user data and channel state information (CSI)

exchange, the high signaling overhead results in practical implementation limitations.

Recently, multiple-input multiple-output (MIMO) transmission with large-scale antenna arrays at the BS has attracted substantial interest from both academia and industry. Using simple linear processing, such large-scale antenna arrays were proved to be able to substantially reduce the effects of the uncorrelated noise, small-scale fading and intracell interference [11, 12]. Then, the energy and spectral efficiency of very large multiuser MIMO uplink systems were investigated in [13], which showed that the power radiated by each terminal could be made inversely proportional to either the number of BS antennas or at least its square-root, considering both perfect and imperfect CSI. In [6], it was stated that the potential benefits have elevated large-scale MIMO to a central position as a promising technology for the next generation of wireless systems. In a HetNet setting, [14, 15] proposed to use large scale antenna arrays at the BS and limited antennas at the SCs¹ due to their smaller form factor. As variable structures of antenna arrays, such as a cylindrical array, requires less space [16], large-scale antenna arrays set at SCs becomes realizable. Recently, NEC Corporation announced that it has developed a prototype of A4-sized massive-element Active Antenna System for 5G small cell base stations, and it proposed the use of a massive-element antenna in small cells for capacity enhancement [17]. Up to date, few papers in the literature have studied the effect of employing massive MIMO at SCs. In [18], *HetNet with large-scale antenna arrays* was investigated on downlink performance with interference coordination, and random matrix theory was used to simplify the analysis significantly. It was shown in [1] that using large-scale antenna arrays in SC reduces both intra-tier interference and the cross-tier interference from other nodes in the HetNet system, leading to higher spectral efficiency and better coverage, especially for hot zones.

This paper presents a comprehensive study of a two-tier network with large-scale antenna arrays set at both BS and SCs. In our preliminary literature [1], maximum-ratio transmission (MRT) precoding was employed based on the estimated channels obtained from the orthogonal training scheme, and downlink capacity lower bounds for a user in the MC and for a user in an SC were derived in closed-form expressions. However, there are still many critical yet unsolved problems. This paper makes the following contributions to address the remaining issues.

¹In this paper, we use SC to denote the SC node and BS to denote the macro cell node for simplicity.

Y. Dai and X. Dong are with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, V8P 5C2, Canada (Email: yongyu@uvic.ca, xdong@ece.uvic.ca). H. Lin is with the Department of Electrical and Information Systems, Osaka Prefecture University, Sakai, Osaka, 599-8531, Japan (Email: lin@eis.osakafu-u.ac.jp).

Part of this work has been presented at the IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, Dec. 2015 [1].

- 1) It was stated in [19] that pilot overhead is proportional to the number of user equipment (UE) for the conventional orthogonal training scheme, i.e., the system performance will degrade as the UE number grows due to heavy pilot overhead. In [11, 12], the pilot reuse (PR) technique is utilized among the macro cells to reduce the pilot overhead, while UEs within a cell use orthogonal pilots. [20] studies pilot reuse in a dense small cell network. In a two-tier HetNet with multiple small cells, massive antenna arrays and large number of UEs, we propose to apply pilot reuse among the SCs in this paper, i.e., the same set of orthogonal pilots is reused among the small cells in one macro-cell. Thus the number of orthogonal pilots is smaller than the total UE number in the whole network.
- 2) We present for the first time the downlink capacity lower bounds of the large-scale HetNet system, where simple linear precoding such as MRT or zero-forcing transmission (ZFT) is employed at each node, followed by detailed asymptotic analysis.
- 3) The design of an efficient and practical user scheduler for the large-scale HetNet is an important and challenging problem, because the required CSI exchange becomes prohibitively complicated due to the large-scale antenna arrays and the large number of UEs in the MC and SCs. Based on the obtained capacity bounds and asymptotic analysis, a greedy scheduling algorithm (GSA) and an asymptotic scheduling algorithm (ASA) are proposed, respectively, where GSA requires only statistical CSI (SCSI) shared between BS and SCs, and ASA even removes the need for any CSI exchange among nodes.

The rest of the paper is organized as follows. We briefly describe the system model for HetNet with large-scale antenna arrays in Section II. In Section III, lower bounds for the achievable rate are derived with both imperfect CSI based MRT and ZFT, followed by corresponding asymptotic analysis. Then, two user scheduling algorithms are developed in Section IV. Moreover, simulation results under different system configurations are given in Section V to demonstrate the effectiveness of both the derived rate expressions and the developed schemes. Finally, conclusions are drawn in Section VI.

Notations: For a matrix \mathbf{X} , we use \mathbf{X}^T , \mathbf{X}^H , \mathbf{X}^* and $\text{Tr}\{\mathbf{X}\}$ to denote the transpose, the Hermitian transpose, the conjugate, and the trace, respectively. \mathbf{I}_N is an $N \times N$ identity matrix and $\mathbf{0}_{M \times N}$ is an $M \times N$ zero matrix. Moreover, $\mathbb{E}[\cdot]$ denotes the expectation operator. The symbol $\|\mathbf{x}\|$ indicates the 2-norm of vector \mathbf{x} , and $\text{diag}\{\mathbf{x}\}$ denotes a diagonal matrix with \mathbf{x} being its diagonal entries. $|x|$ is the absolute value of x , while $|U|$ represents the number of elements in set U . Finally, the notation $\xrightarrow{a.s.}$ means almost sure convergence, and $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{D}_x)$ represents a circularly symmetric complex Gaussian vector \mathbf{x} with zero mean and covariance matrix \mathbf{D}_x .

II. SYSTEM MODEL

Fig. 1 shows the considered two-tier network architecture with one cell consisting of one macro BS, which is overlaid

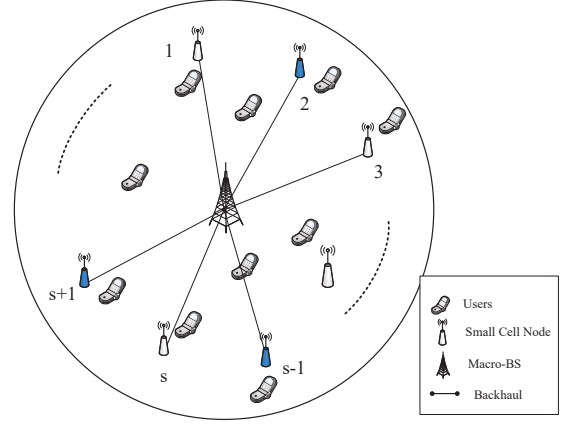


Fig. 1: System model for HetNet with SCs deployment.

with a dense tier of S uniformly distributed SCs by sharing the same time-frequency resources. Assume that the BS and SCs are respectively equipped with large-scale arrays of N_{BS} and N_{SC} antennas, where $N_{BS} > N_{SC} \gg 1$, while each user has only one antenna due to the size or complexity constraint. Notably, uniform user distribution in the cell is focused here. Based on the biased user association [3], the users served by the macro BS are designated to a macro UE (MUE) set, and those served by each SC are designated to a small cell UE (SUE) set. Furthermore, suppose that the macro BS serves K MUEs simultaneously while each SC serves L SUEs with $K \leq N_{BS}$ and $L \leq N_{SC}$. Denote the MUE and SUE sets as U_M and $U_S^{(m)}$, respectively, then we have $K \leq |U_M|$ and $L \leq |U_S^{(m)}|$. The selected subsets of MUE and SUE after user scheduling are denoted by I and J_m ($m \in \{1, \dots, S\}$), respectively.

For the channel matrices, they account for both small-scale fading and large-scale fading. Here, we assume that all the channels between the users and the nodes follow independent and identically distributed (i.i.d.) Rayleigh fading and time division duplex (TDD) is adopted with channel reciprocity satisfied. Denote the channel matrices from the BS and n th ($n \in \{1, \dots, S\}$) SC to the K MUEs as $\mathbf{G}_{B-M} = [\mathbf{g}_{B-M}^{(1)}, \dots, \mathbf{g}_{B-M}^{(K)}] \in \mathbb{C}^{N_{BS} \times K}$ and $\mathbf{G}_{S-M}^{(n)} = [\mathbf{g}_{S-M}^{(n,1)}, \dots, \mathbf{g}_{S-M}^{(n,K)}] \in \mathbb{C}^{N_{SC} \times K}$, respectively, and use $\mathbf{G}_{B-S}^{(m)} = [\mathbf{g}_{B-S}^{(m,1)}, \dots, \mathbf{g}_{B-S}^{(m,L)}] \in \mathbb{C}^{N_{BS} \times L}$ and $\mathbf{G}_{S-S}^{(n,m)} = [\mathbf{g}_{S-S}^{(n,m,1)}, \dots, \mathbf{g}_{S-S}^{(n,m,L)}] \in \mathbb{C}^{N_{SC} \times L}$ to represent the channel matrices from the BS and the n th SC to the L SUEs in the m th SC, respectively. We have $\mathbf{G}_{B-M} = \mathbf{H}_{B-M} \mathbf{D}_{B-M}^{1/2}$, $\mathbf{G}_{B-S}^{(m)} = \mathbf{H}_{B-S}^{(m)} (\mathbf{D}_{B-S}^{(m)})^{1/2}$, $\mathbf{G}_{S-M}^{(n)} = \mathbf{H}_{S-M}^{(n)} (\mathbf{D}_{S-M}^{(n)})^{1/2}$ and $\mathbf{G}_{S-S}^{(n,m)} = \mathbf{H}_{S-S}^{(n,m)} (\mathbf{D}_{S-S}^{(n,m)})^{1/2}$ where $n, m \in \{1, \dots, S\}$, the first items $\mathbf{H}_{B-M} \in \mathbb{C}^{N_{BS} \times K}$, $\mathbf{H}_{B-S}^{(m)} \in \mathbb{C}^{N_{BS} \times L}$, $\mathbf{H}_{S-M}^{(n)} \in \mathbb{C}^{N_{SC} \times K}$ and $\mathbf{H}_{S-S}^{(n,m)} \in \mathbb{C}^{N_{SC} \times L}$ include the i.i.d. $\mathcal{CN}(0, 1)$ small-scale fading coefficients, and the second items are the large-scale fading diagonal

matrices given by $\mathbf{D}_{B-M} = \text{diag}\{\beta_{B-M}^{(1)}, \dots, \beta_{B-M}^{(K)}\}$, $\mathbf{D}_{B-S}^{(m)} = \text{diag}\{\beta_{B-S}^{(m,1)}, \dots, \beta_{B-S}^{(m,L)}\}$, $\mathbf{D}_{S-M}^{(n)} = \text{diag}\{\beta_{S-M}^{(n,1)}, \dots, \beta_{S-M}^{(n,K)}\}$, and $\mathbf{D}_{S-S}^{(n,m)} = \text{diag}\{\beta_{S-S}^{(n,m,1)}, \dots, \beta_{S-S}^{(n,m,L)}\}$.

A. Channel Estimation with Pilot Reuse

Practically, the channel matrix from each node to its corresponding users, i.e., \mathbf{G}_{B-M} and $\mathbf{G}_{S-S}^{(m,m)}$ ($m \in \{1, \dots, S\}$), have to be estimated based on the uplink training. At the beginning of each coherence interval T , all users simultaneously transmit pilot sequences of length τ symbols. On account of the slight interferences between low-power SCs which are far away from each other, we present a pilot reuse pattern for small cells in a large-scale HetNet system.

First, we denote the reuse factor as γ , i.e., all SCs utilize γ sets of L pairwise orthogonal pilot sequences with a total of $SS = S/\gamma$ SCs sharing the same set. This requires $\tau \geq K + L \times \gamma$ to satisfy the orthogonality of the MC and SC pilot sets. Then, we group SCs into γ sets according to the maximum relative distance criterion and SCs in one set use the same pilot sequences. Since all low-power nodes are modeled as uniformly distributed in a circle with BS at the center as shown in Fig. 1, we can denote the n th ($n \in \{1, \dots, \gamma\}$) SC set as $\mathcal{A}_n = \{n, n + \gamma, \dots, n + (SS - 1)\gamma\}$. Taking $S = 8$ and reuse factor $\gamma = 2$ for example, the SC sets are $\mathcal{A}_1 = \{1, 3, 5, 7\}$ and $\mathcal{A}_2 = \{2, 4, 6, 8\}$. The 4 SCs in each set share one pilot set which includes L pairwise orthogonal pilot sequences, and there are 2 pilot sets for the total of 8 SCs.

Then the training matrix received at the BS and the m th ($m \in \mathcal{A}_r$) SC can be written as

$$\begin{aligned} \mathbf{Y}_{BS} &= \sqrt{\tau p_\tau} \left(\mathbf{G}_{B-M} \Phi_{MUE} + \sum_{t=1}^{\gamma} \sum_{l \in \mathcal{A}_t} \mathbf{G}_{B-S}^{(l)} \Phi_{SUE}^{(t)} \right) + \mathbf{N}_{BS} \\ \mathbf{Y}_{SC}^{(m)} &= \sqrt{\tau p_\tau} \left(\sum_{t=1}^{\gamma} \sum_{l \in \mathcal{A}_t} \mathbf{G}_{S-S}^{(m,l)} \Phi_{SUE}^{(t)} + \mathbf{G}_{S-M}^{(m)} \Phi_{MUE} \right) + \mathbf{N}_{SC}^{(m)} \end{aligned} \quad (1)$$

respectively, where p_τ is the transmit power of each pilot symbol, \mathbf{N}_{BS} and $\mathbf{N}_{SC}^{(m)}$ are the additive white Gaussian noise (AWGN) matrices with i.i.d. components following $\mathcal{CN}(0, \sigma_0^2)$, the training vectors transmitted by the i th ($i \in I$) MUE is denoted by the i th row of $\Phi_{MUE} \in \mathbb{C}^{K \times \tau}$, satisfying $\Phi_{MUE} \Phi_{MUE}^H = \mathbf{I}_K$, while the training vector transmitted by the j th ($j \in J_m$) SUE of one SC in the r th set \mathcal{A}_r is represented by the j th row of $\Phi_{SUE}^{(r)} \in \mathbb{C}^{L \times \tau}$, satisfying $\Phi_{SUE}^{(r)} (\Phi_{SUE}^{(r)})^H = \mathbf{I}_L$. Moreover, since the rows of pilot sequence matrices are pairwise orthogonal, we have $\Phi_{MUE} (\Phi_{SUE}^{(r)})^H = \mathbf{0}_{K \times L}$ and $\Phi_{SUE}^{(r)} (\Phi_{SUE}^{(t)})^H = \mathbf{0}_{L \times L}$ ($\forall r \neq t \in \{1, \dots, \gamma\}$).

In order to estimate \mathbf{G}_{B-M} and $\mathbf{G}_{S-S}^{(m,m)}$ ($m \in \mathcal{A}_r$), we employ the minimum mean-square-error (MMSE) estimation

at each node [21]. The estimated channels are given by

$$\begin{aligned} \hat{\mathbf{G}}_{B-M} &= \frac{1}{\sqrt{\tau p_\tau}} \mathbf{Y}_{BS} \Phi_{MUE}^H \tilde{\mathbf{D}}_{B-M} \\ &= \mathbf{G}_{B-M} \tilde{\mathbf{D}}_{B-M} + \frac{1}{\sqrt{\tau p_\tau}} \tilde{\mathbf{N}}_{BS} \tilde{\mathbf{D}}_{B-M} \\ \hat{\mathbf{G}}_{S-S}^{(m,m)} &= \frac{1}{\sqrt{\tau p_\tau}} \mathbf{Y}_{SC}^{(m)} (\Phi_{SUE}^{(r)})^H \tilde{\mathbf{D}}_{S-S}^{(m,m)} \\ &= \sum_{l \in \mathcal{A}_r} \mathbf{G}_{S-S}^{(m,l)} \tilde{\mathbf{D}}_{S-S}^{(m,m)} + \frac{1}{\sqrt{\tau p_\tau}} \tilde{\mathbf{N}}_{SC}^{(m)} \tilde{\mathbf{D}}_{S-S}^{(m,m)} \end{aligned} \quad (2)$$

where $\tilde{\mathbf{D}}_{B-M} \triangleq \left(\frac{\mathbf{D}_{B-M}^{-1} \sigma_0^2}{\tau p_\tau} + \mathbf{I}_K \right)^{-1}$, $\tilde{\mathbf{D}}_{S-S}^{(m,m)} \triangleq \left[\left(\sum_{l \neq m, l \in \mathcal{A}_r} \mathbf{D}_{S-S}^{(m,l)} + \frac{\sigma_0^2}{\tau p_\tau} \mathbf{I}_L \right) (\mathbf{D}_{S-S}^{(m,m)})^{-1} + \mathbf{I}_L \right]^{-1}$, $\tilde{\mathbf{N}}_{BS} \triangleq \mathbf{N}_{BS} \Phi_{MUE}^H$ and $\tilde{\mathbf{N}}_{SC}^{(m)} \triangleq \mathbf{N}_{SC}^{(m)} (\Phi_{SUE}^{(r)})^H$ ($m \in \mathcal{A}_r$) are defined. Due to the property of Φ_{MUE} and $\Phi_{SUE}^{(r)}$, $\tilde{\mathbf{N}}_{BS}$ and $\tilde{\mathbf{N}}_{SC}^{(m)}$ are also composed of i.i.d. $\mathcal{CN}(0, \sigma_0^2)$ elements. Then, we have

$$\mathbf{G}_{B-M} = \hat{\mathbf{G}}_{B-M} + \Xi_{B-M}, \mathbf{G}_{S-S}^{(m,m)} = \hat{\mathbf{G}}_{S-S}^{(m,m)} + \Xi_{S-S}^{(m,m)} \quad (3)$$

where Ξ_{B-M} and $\Xi_{S-S}^{(m,m)}$ denote the estimation error matrices which are independent of $\hat{\mathbf{G}}_{B-M}$ and $\hat{\mathbf{G}}_{S-S}^{(m,m)}$ from the property of MMSE channel estimation [21]. Hence, we have $\hat{\mathbf{G}}_{B-M} \sim \mathcal{CN}(0, \hat{\mathbf{D}}_{B-M})$ with $\hat{\mathbf{D}}_{B-M} = \text{diag}\{\hat{\beta}_{B-M}^{(1)}, \dots, \hat{\beta}_{B-M}^{(K)}\}$, $\hat{\mathbf{G}}_{S-S}^{(m,m)} \sim \mathcal{CN}(0, \hat{\mathbf{D}}_{S-S}^{(m,m)})$ with $\hat{\mathbf{D}}_{S-S}^{(m,m)} = \text{diag}\{\hat{\beta}_{S-S}^{(m,m,1)}, \dots, \hat{\beta}_{S-S}^{(m,m,L)}\}$, $\Xi_{B-M} \sim \mathcal{CN}(0, \mathbf{D}_{B-M} - \hat{\mathbf{D}}_{B-M})$ with the i th column vector denoted by $\xi_{B-M}^{(i)}$, and $\Xi_{S-S}^{(m,m)} \sim \mathcal{CN}(0, \mathbf{D}_{S-S}^{(m,m)} - \hat{\mathbf{D}}_{S-S}^{(m,m)})$ with the j th column vector denoted by $\xi_{S-S}^{(m,m,j)}$. Here, the estimated

large-scale fading factors satisfy $\hat{\beta}_{B-M}^{(i)} = \frac{\tau p_\tau (\beta_{B-M}^{(i)})^2}{\tau p_\tau \beta_{B-M}^{(i)} + \sigma_0^2}$, $i \in I$ and $\hat{\beta}_{S-S}^{(m,m,j)} = \frac{\tau p_\tau (\beta_{S-S}^{(m,m,j)})^2}{\tau p_\tau \sum_{l \in \mathcal{A}_r} \beta_{S-S}^{(m,l,j)} + \sigma_0^2}$, where $j \in J_m$ and $m \in \mathcal{A}_r$.

B. Data Transmission

In the downlinks, the received signals at K MUEs and L SUEs in the m th small cell are

$$\begin{aligned} \mathbf{y}_M &= \mathbf{G}_{B-M}^T \mathbf{W}_{BS} \mathbf{x}_{BS} + \sum_{n=1}^S (\mathbf{G}_{S-M}^{(n)})^T \mathbf{W}_{SC}^{(n)} \mathbf{x}_{SC}^{(n)} + \mathbf{n}_M \\ \mathbf{y}_S^{(m)} &= (\mathbf{G}_{B-S}^{(m)})^T \mathbf{W}_{BS} \mathbf{x}_{BS} + \sum_{n=1}^S (\mathbf{G}_{S-S}^{(m,n)})^T \mathbf{W}_{SC}^{(n)} \mathbf{x}_{SC}^{(n)} + \mathbf{n}_S^{(m)} \end{aligned} \quad (4)$$

respectively, where \mathbf{W}_{BS} and $\mathbf{W}_{SC}^{(n)}$ represent the linear precoding matrices at the BS and n th SC, respectively; $\mathbf{x}_{BS} = [x_{BS}^{(1)}, \dots, x_{BS}^{(K)}]^T$ and $\mathbf{x}_{SC}^{(n)} = [x_{SC}^{(n,1)}, \dots, x_{SC}^{(n,L)}]^T$ are the complex-valued data symbols from BS to its MUEs and from n th SC to its own SUEs, respectively, satisfying

$\mathbb{E}[\mathbf{x}_{\text{BS}}\mathbf{x}_{\text{BS}}^H] = \mathbf{I}_{N_{\text{BS}}}$ and $\mathbb{E}[\mathbf{x}_{\text{SC}}^{(n)}(\mathbf{x}_{\text{SC}}^{(n)})^H] = \mathbf{I}_{N_{\text{SC}}}$; and $\mathbf{n}_{\text{M}} = [n_{\text{M}}^{(1)}, \dots, n_{\text{M}}^{(K)}]^T$ and $\mathbf{n}_{\text{S}}^{(m)} = [n_{\text{S}}^{(m,1)}, \dots, n_{\text{S}}^{(m,L)}]^T$ involves the AWGN of variance σ_0^2 .

C. MRT Precoding

Aiming to maximize the received signal-to-noise ratio, the MRT technique is utilized at both the BS and SCs to process the transmit signals towards the corresponding users. Given the estimated channel state information, the MRT precoding is expressed as [22]

$$\mathbf{W}_{\text{BS}} = \alpha_{\text{BS}} \hat{\mathbf{G}}_{\text{B-M}}^*, \quad \mathbf{W}_{\text{SC}}^{(m)} = \alpha_{\text{SC}}^{(m)} \left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^* \quad (6)$$

where α_{BS} and $\alpha_{\text{SC}}^{(m)}$ are normalization constants, chosen to satisfy the transmit power constraints at the BS and SCs, respectively. On the basis of (6) and $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$, we have

$$\alpha_{\text{BS}} = \sqrt{\frac{p_{\text{BS}}}{N_{\text{BS}}\Phi_{\text{B-M}}}}, \quad \alpha_{\text{SC}}^{(m)} = \sqrt{\frac{p_{\text{SC}}^{(m)}}{N_{\text{SC}}\Phi_{\text{S-S}}^{(m)}}} \quad (7)$$

where $\Phi_{\text{B-M}} = \sum_{i=1}^K \hat{\beta}_{\text{B-M}}^{(i)}$, and $\Phi_{\text{S-S}}^{(m)} = \sum_{l=1}^L \hat{\beta}_{\text{S-S}}^{(m,m,l)}$ with $m \in \{1, \dots, S\}$.

D. ZFT Precoding

Likewise, when ZFT is employed based on imperfect CSI, in which the pseudo-inverse of the estimated channels in (3) are utilized for linear precoding, the precoder is given by [22]

$$\begin{aligned} \mathbf{W}_{\text{BS}} &= \alpha_{\text{BS}} \hat{\mathbf{G}}_{\text{B-M}}^* \left(\hat{\mathbf{G}}_{\text{B-M}}^T \hat{\mathbf{G}}_{\text{B-M}}^* \right)^{-1} = \alpha_{\text{BS}} \hat{\mathbf{G}}_{\text{B-M}}^* \\ \mathbf{W}_{\text{SC}}^{(m)} &= \alpha_{\text{SC}}^{(m)} \left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^* \left[\left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^T \left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^* \right]^{-1} \\ &= \alpha_{\text{SC}}^{(m)} \left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^* \end{aligned} \quad (8)$$

where $\hat{\mathbf{G}}_{\text{B-M}} = \hat{\mathbf{G}}_{\text{B-M}} \left(\hat{\mathbf{G}}_{\text{B-M}}^H \hat{\mathbf{G}}_{\text{B-M}} \right)^{-1}$, $\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} = \hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \left[\left(\hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right)^H \hat{\mathbf{G}}_{\text{S-S}}^{(m,m)} \right]^{-1}$, α_{BS} and $\alpha_{\text{SC}}^{(m)}$ are normalization constants. Similarly, based on (8) and $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$, we have

$$\alpha_{\text{BS}} = \sqrt{\frac{(N_{\text{BS}} - K - 1)p_{\text{BS}}}{\Psi_{\text{B-M}}}}, \quad \alpha_{\text{SC}}^{(m)} = \sqrt{\frac{(N_{\text{SC}} - L - 1)p_{\text{SC}}^{(m)}}{\Psi_{\text{S-S}}^{(m)}}} \quad (9)$$

where $\Psi_{\text{B-M}} = \sum_{i=1}^K \frac{1}{\hat{\beta}_{\text{B-M}}^{(i)}}$, and $\Psi_{\text{S-S}}^{(m)} = \sum_{l=1}^L \frac{1}{\hat{\beta}_{\text{S-S}}^{(m,m,l)}}$ with $m \in \{1, \dots, S\}$. The detailed derivation is given in Appendix VII.

III. ACHIEVABLE RATE ANALYSIS

The exact rate analysis of the MUE and SUE in the pilot assisted massive MIMO heterogeneous network considered is highly complicated and intractable. In this section, we

provide a closed-form capacity lower bound of each user for both MRT and ZFT precoding, respectively. The simple lower bounds can be applied to user scheduling and power allocation optimization as detailed in subsequent sections.

A. MRT Precoding

In practice, only imperfect CSI derived from transmitted pilots is available at each node for linear precoding. Utilizing MRT precoding in (4) and (5), the received signal of the i th ($i \in I$) MUE and j th ($j \in J_m$) SUE at the m th ($m \in \{1, \dots, S\}$) SC can be rewritten as (10) and (11). Note that both the BS and small cell nodes treat the estimated channels as the true channels [13], and the first term is the desired signal. The remaining terms are considered as interferences and noise, including estimation error caused interference term. Accordingly, with imperfect CSI, the ergodic achievable rate of MUE i ($i \in I$) and SUE j ($j \in J_m$) in the m th ($m \in \{1, \dots, S\}$) SC are given by

$$R_{\text{M}}^{(i)} = \mathbb{E} \left[\log_2 \left(1 + \frac{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4}{\text{EEI}_i + \text{IMI}_i + \text{CTI}_i + \sigma_0^2} \right) \right] \quad (12)$$

$$R_{\text{S}}^{(m,j)} = \mathbb{E} \left[\log_2 \left(1 + \frac{(\alpha_{\text{SC}}^{(m)})^2 \|\hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)}\|^4}{\text{EEI}_{m,j} + \text{CTI}_{m,j} + \text{ISI}_{m,j} + \text{SSI}_{m,j} + \sigma_0^2} \right) \right] \quad (13)$$

respectively, where EEI_i , IMI_i and CTI_i denote the estimation error induced interference, the intra-MC interference and the cross-tier interference for the i th MUE, respectively, given by

$$\begin{aligned} \text{EEI}_i &= \alpha_{\text{BS}}^2 \left| \left(\hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right|^2, \quad \text{IMI}_i = \sum_{k \neq i}^K \alpha_{\text{BS}}^2 \left| \left(\hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(k)} \right|^2 \\ \text{CTI}_i &= \sum_{n=1}^S \sum_{l=1}^L (\alpha_{\text{SC}}^{(n)})^2 \left| \left(\hat{\mathbf{g}}_{\text{S-M}}^{(n,i)} \right)^H \hat{\mathbf{g}}_{\text{S-M}}^{(n,i)} \right|^2, \end{aligned} \quad (14)$$

and $\text{EEI}_{m,j}$, $\text{CTI}_{m,j}$, $\text{ISI}_{m,j}$ and $\text{SSI}_{m,j}$ denote the estimation error induced interference, the cross-tier interference, the intra-SC interference and the inter-SC interference for the j th SUE in the m th SC, respectively, given by

$$\begin{aligned} \text{EEI}_{m,j} &= (\alpha_{\text{SC}}^{(m)})^2 \left| \left(\hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right|^2 \\ \text{CTI}_{m,j} &= \sum_{i=1}^K \alpha_{\text{BS}}^2 \left| \left(\hat{\mathbf{g}}_{\text{B-S}}^{(m,j)} \right)^H \hat{\mathbf{g}}_{\text{B-S}}^{(i)} \right|^2 \\ \text{ISI}_{m,j} &= \sum_{l_1 \neq j}^L (\alpha_{\text{SC}}^{(m)})^2 \left| \left(\hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,l_1)} \right|^2 \\ \text{SSI}_{m,j} &= \sum_{n \neq m}^S \sum_{l_2=1}^L (\alpha_{\text{SC}}^{(n)})^2 \left| \left(\hat{\mathbf{g}}_{\text{S-S}}^{(n,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(n,m,l_2)} \right|^2. \end{aligned} \quad (15)$$

Notably, the inter-SC interference $\text{SSI}_{m,j}$ includes the pilot contamination effect caused by pilot reuse. In the above

$$y_M^{(i)} = \underbrace{\alpha_{BS} \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^* x_{BS}^{(i)}}_{\text{desired signal}} + \underbrace{\alpha_{BS} \left(\xi_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^* x_{BS}^{(i)}}_{\text{estimation error induced interference}} + \underbrace{\sum_{k \neq i}^K \alpha_{BS} \left(\mathbf{g}_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(k)} \right)^* x_{BS}^{(k)}}_{\text{intra-MC interference}} \quad (10)$$

$$+ \underbrace{\sum_{n=1}^S \sum_{l=1}^L \alpha_{SC}^{(n)} \left(\mathbf{g}_{S-M}^{(n,i)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(n,n,l)} \right)^* x_{SC}^{(n,l)}}_{\text{cross-tier interference}} + \underbrace{n_M^{(i)}}_{\text{noise at MUE}}$$

$$y_S^{(m,j)} = \underbrace{\alpha_{SC}^{(m)} \left(\hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right)^* x_{SC}^{(m,j)}}_{\text{desired signal}} + \underbrace{\alpha_{SC}^{(m)} \left(\xi_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right)^* x_{SC}^{(m,j)}}_{\text{estimation error induced interference}} + \underbrace{\sum_{i=1}^K \alpha_{BS} \left(\mathbf{g}_{B-S}^{(m,j)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^* x_{BS}^{(i)}}_{\text{cross-tier interference}} \\ + \underbrace{\sum_{l_1 \neq j}^L \alpha_{SC}^{(m)} \left(\mathbf{g}_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,l_1)} \right)^* x_{SC}^{(m,l_1)}}_{\text{intra-SC interference}} + \underbrace{\sum_{n \neq m}^S \sum_{l_2=1}^L \alpha_{SC}^{(n)} \left(\mathbf{g}_{S-S}^{(n,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(n,n,l_2)} \right)^* x_{SC}^{(n,l_2)}}_{\text{inter-SC interference}} + \underbrace{n_S^{(m,j)}}_{\text{noise at SUE}}. \quad (11)$$

achievable rate expressions, expectations over the estimated instantaneous CSI cannot be further derived into tractable forms. Therefore, we adopt a similar bounding technique of [13] to obtain closed form rate expressions, the result of which will provide insights on the impact of different system parameters and facilitate further optimizations.

By the convexity of $\log_2(1 + \frac{1}{x})$ and Jensen's inequality, from (12) and (13), a lower bound on the achievable rate is obtained as

$$R_{0,M,M}^{(i)} = \log_2 \left(1 + \left(\mathbb{E} \left[\frac{\text{EEI}_i + \text{IMI}_i + \text{CTI}_i + \sigma_0^2}{\alpha_{BS}^2 \left\| \hat{\mathbf{g}}_{B-M}^{(i)} \right\|^4} \right] \right)^{-1} \right) \quad (16)$$

$$R_{0,S,M}^{(m,j)} = \log_2 \left(1 + \left(\mathbb{E} \left[\frac{\text{EEI}_{m,j} + \text{CTI}_{m,j} + \text{ISI}_{m,j} + \text{SSI}_{m,j} + \sigma_0^2}{\left(\alpha_{SC}^{(m)} \right)^2 \left\| \hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right\|^4} \right] \right)^{-1} \right). \quad (17)$$

Theorem 1: With imperfect CSI based MRT, $N_{BS} \geq 2$ and $N_{SC} \geq 2$, the downlink achievable rate of the i th ($i \in I$) MUE and j th ($j \in J_m$) SUE in the m th ($m \in \{1, \dots, S\}$) SC, for finite N_{BS} and N_{SC} , are lower bounded by

$$R_{0,M,M}^{(i)} = \log_2 \left(1 + \frac{a_{MR}^{(i)} p_{BS}}{b_{MR}^{(i)} p_{BS} + \sum_{n=1}^S c_{MR}^{(n,i)} p_{SC}^{(n)} + \sigma_0^2} \right) \quad (18)$$

$$R_{0,S,M}^{(m,j)} = \log_2 \left(1 + \frac{d_{MR}^{(m,j)} p_{SC}^{(m)}}{\sum_{n=1}^S e_{MR}^{(n,m,j)} p_{SC}^{(n)} + f_{MR}^{(m,j)} p_{BS} + \sigma_0^2} \right) \quad (19)$$

$$\text{where } a_M^{(i)} = \frac{(N_{BS}-1)(N_{BS}-2)(\hat{\beta}_{B-M}^{(i)})^2}{N_{BS}\Phi_{B-M}}, \quad b_M^{(i)} = \beta_{B-M}^{(i)} - \frac{2}{N_{BS}}\hat{\beta}_{B-M}^{(i)} - \frac{(N_{BS}-4)(\hat{\beta}_{B-M}^{(i)})^2 + 2\beta_{B-M}^{(i)}\hat{\beta}_{B-M}^{(i)}}{N_{BS}\Phi_{B-M}}, \quad c_M^{(n,i)} = \beta_{S-M}^{(n,i)}, \\ d_M^{(m,j)} = \frac{(N_{SC}-1)(N_{SC}-2)(\hat{\beta}_{S-S}^{(m,m,j)})^2}{N_{SC}\Phi_{S-S}}, \quad e_M^{(n,m,j)} = \begin{cases} \beta_{S-S}^{(n,m,j)}, & n \neq m, n \notin \mathcal{A}_r \\ \frac{N_{SC}(\beta_{S-S}^{(n,m,j)})^2(\hat{\beta}_{S-S}^{(n,n,j)})^2}{(\beta_{S-S}^{(n,n,j)})^2\Phi_{S-S}^{(n)}} + \beta_{S-S}^{(n,m,j)}, & n \neq m, n \in \mathcal{A}_r \\ \beta_{S-S}^{(m,m,j)} - \frac{2}{N_{SC}}\hat{\beta}_{S-S}^{(m,m,j)} - \frac{(N_{SC}-4)(\hat{\beta}_{S-S}^{(m,m,j)})^2 + 2\beta_{S-S}^{(m,m,j)}\hat{\beta}_{S-S}^{(m,m,j)}}{N_{SC}\Phi_{S-S}^{(m)}}, & n = m \end{cases} \\ \text{and } f_M^{(m,j)} = \beta_{B-S}^{(m,j)}.$$

Proof: See Appendix VIII.

Remark 1: Since the proof in Appendix VIII does not use any asymptotic assumptions on the antenna size, Theorem 1 is also valid for conventional scale MIMO systems. The capacity lower bounds for perfect CSI can be obtained by setting $\hat{\beta}_{\cdot}^{(\cdot)} = \beta_{\cdot}^{(\cdot)}$ in (18) and (19). Moreover, it can be observed from Appendix VIII that all the interferences contained in the received signals of the MUEs (i.e., EEI, IMI and CTI) can be significantly mitigated relative to the desired signals by increasing N_{BS} . Similarly, all the interference effect at the SUEs (i.e., EEI, CTI, ISI and SSI) is able to be reduced by increasing N_{SC} . These observations support the use of large scale antenna arrays at both BS and SC. In addition, the expression of $R_{0,M,M}^{(i)}$ indicates that the MUE rate increases monotonically with N_{BS} but has no relationship with N_{SC} . Similarly, the SUE rate increases monotonically with N_{SC} and is independent of N_{BS} .

B. ZFT Precoding

For imperfect CSI based ZFT precoding, the received signal can be rewritten as (20) and (21), respectively, where $(\hat{\mathbf{g}}_{B-M}^{(i)})^T (\hat{\mathbf{g}}_{B-M}^{(i)})^* = 1$, $(\hat{\mathbf{g}}_{S-S}^{(m,m,j)})^T (\hat{\mathbf{g}}_{S-S}^{(m,m,j)})^* = 1$, and the intra-MC interference IMI_i and intra-SC interference

$$y_M^{(i)} = \underbrace{\alpha_{BS} x_{BS}^{(i)}}_{\text{desired signal}} + \underbrace{\alpha_{BS} \left(\xi_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^* x_{BS}^{(i)}}_{\text{estimation error induced interference}} + \underbrace{\sum_{k \neq i} \alpha_{BS} \left(\xi_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(k)} \right)^* x_{BS}^{(k)}}_{\text{intra-MC interference}} \quad (20)$$

$$+ \underbrace{\sum_{n=1}^S \sum_{l=1}^L \alpha_{SC}^{(n)} \left(\mathbf{g}_{S-M}^{(n,i)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(n,l)} \right)^* x_{SC}^{(n,l)}}_{\text{cross-tier interference}} + \underbrace{n_M^{(i)}}_{\text{noise at MUE}}$$

$$y_S^{(m,j)} = \underbrace{\alpha_{SC}^{(m)} x_{SC}^{(m,j)}}_{\text{desired signal}} + \underbrace{\alpha_{SC}^{(m)} \left(\xi_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right)^* x_{SC}^{(m,j)}}_{\text{estimation error induced interference}} + \underbrace{\sum_{i=1}^K \alpha_{BS} \left(\mathbf{g}_{B-S}^{(m,j)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^* x_{BS}^{(i)}}_{\text{cross-tier interference}} \quad (21)$$

$$+ \underbrace{\sum_{l_1 \neq j}^L \alpha_{SC}^{(m)} \left(\xi_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,l_1)} \right)^* x_{SC}^{(m,l_1)}}_{\text{intra-SC interference}} + \underbrace{\sum_{n \neq m}^S \sum_{l_2=1}^L \alpha_{SC}^{(n)} \left(\mathbf{g}_{S-S}^{(n,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(n,n,l_2)} \right)^* x_{SC}^{(n,l_2)}}_{\text{inter-SC interference}} + \underbrace{n_S^{(m,j)}}_{\text{noise at SUE}}$$

$ISI_{m,j}$ are reduced because ZFT precoding is able to null multi-user interference signals, i.e., $\left(\hat{\mathbf{g}}_{B-M}^{(i)} \right)^T \left(\hat{\mathbf{g}}_{B-M}^{(k)} \right)^* = 0$ and $\left(\hat{\mathbf{g}}_{S-S}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{S-S}^{(m,m,l)} \right)^* = 0$ for $\forall k \neq i$ and $\forall l \neq j$. Similarly, by the convexity of $\log_2 \left(1 + \frac{1}{x} \right)$ and Jensen's inequality, capacity lower bounds of MUE i ($i \in I$) and SUE j ($j \in J_m$) in the m th ($m \in \{1, \dots, S\}$) SC in (16) and (17) become

$$R_{0,M,Z}^{(i)} = \log_2 \left(1 + \left(\mathbb{E} \left[\frac{EEL_i + IMI_i + CTI_i + \sigma_0^2}{\alpha_{BS}^2} \right] \right)^{-1} \right) \quad (22)$$

$$R_{0,S,Z}^{(m,j)} = \log_2 \left(1 + \left(\mathbb{E} \left[\frac{EEL_{m,j} + CTI_{m,j} + ISI_{m,j} + SSI_{m,j} + \sigma_0^2}{\left(\alpha_{SC}^{(m)} \right)^2} \right] \right)^{-1} \right) \quad (23)$$

where (24) and (25).

Theorem 2: With imperfect CSI based ZFT, $N_{BS} \geq 2$ and $N_{SC} \geq 2$, the downlink achievable rate of the i th ($i \in I$) MUE and j th ($j \in J_m$) SUE in the m th ($m \in \{1, \dots, S\}$) SC, for finite N_{BS} and N_{SC} , are lower bounded by

$$R_{0,M,Z}^{(i)} = \log_2 \left(1 + \frac{a_{ZF}^{(i)} p_{BS}}{b_{ZF}^{(i)} p_{BS} + \sum_{n=1}^S c_{ZF}^{(n,i)} p_{SC}^{(n)} + \sigma_0^2} \right) \quad (26)$$

$$R_{0,S,Z}^{(m,j)} = \log_2 \left(1 + \frac{d_{ZF}^{(m,j)} p_{SC}^{(m)}}{\sum_{n=1}^S e_{ZF}^{(n,m,j)} p_{SC}^{(n)} + f_{ZF}^{(m,j)} p_{BS} + \sigma_0^2} \right) \quad (27)$$

where $a_Z^{(i)} = \frac{N_{BS}-K-1}{\Psi_{B-M}^{(i)}}$, $b_Z^{(i)} = \xi_{B-M}^{(i)}$, $c_Z^{(n,i)} = \beta_{S-M}^{(n,i)}$, $d_Z^{(m,j)} = \frac{N_{SC}-L-1}{\Psi_{S-S}^{(m)}}$, $f_Z^{(m,j)} = \beta_{B-S}^{(m,j)}$, and $e_Z^{(n,m,j)} =$

$$\begin{cases} \beta_{S-S}^{(n,m,j)}, & n \neq m, n \notin \mathcal{A}_r \\ \beta_{S-S}^{(n,m,j)} - \frac{\beta_{S-S}^{(n,m,j)}}{\Psi_{S-S}^{(n)} \hat{\beta}_{S-S}^{(n,m,j)}} + \frac{(N_{SC}-L-1)(\beta_{S-S}^{(n,m,j)})^2}{\Psi_{S-S}^{(n)} (\beta_{S-S}^{(n,m,j)})^2}, & n \neq m, n \in \mathcal{A}_r \\ \xi_{S-S}^{(m,m,j)}, & n = m. \end{cases}$$

Proof: See Appendix IX.

Remark 2: Similar to MRT, conclusions in Remark 1 are also valid for ZFT based Theorem 2. Moreover, the capacity lower bounds in (26) and (27) indicate that $R_{0,M,ZF}^{(i)}$ decreases monotonically as the estimation error $\xi_{B-M}^{(i)}$ increases for fixed Ψ_{B-M} . Similar conclusions are drawn for the capacity lower bounds of SUEs.

Remark 3: From (26), we can conclude that the expression of $R_{0,M,Z}^{(i)}$ for the i th MUE involves only N_{BS} , but no N_{SC} , which indicates that the capacity lower bound of MUE depends on N_{BS} but has no relationship with N_{SC} . Furthermore, the signal-to-interference-plus-noise ratio (SINR) involved in $R_{0,M,Z}^{(i)}$ is approximately a linearly increasing function of N_{BS} , when $N_{BS} \gg K$. Similarly, from (27), it can be concluded that the capacity lower bound of SUE increases monotonically with N_{SC} but is independent of N_{BS} . However, due to the pilot contamination effect caused by pilot reuse, the increase of the SINR in $R_{0,S,Z}^{(m,j)}$ is not linear even when $N_{SC} \gg L$.

C. Asymptotic Analysis with Massive Arrays

Having obtained the closed-form expressions for the achievable rate in (18) and (19), this subsection provides the asymptotic analysis under two different cases when the number of antennas approaches infinity. Suppose that all SCs have the same transmit power, i.e., $p_{SC}^{(1)} = \dots = p_{SC}^{(S)} = p_{SC}$, and $N_{BS} = \lambda N_{SC}$ with $\lambda \geq 10$.

Proposition 1: In case I where p_τ is fixed, $p_{SC}^{(s)} = p_{SC} = \frac{E_{SC}}{N_{SC}^{xt}}$ ($s = 1, \dots, S$), $p_{BS} = \frac{E_{BS}}{N_{BS}^{xt}}$, and E_{SC} and E_{BS} are fixed, to achieve non-vanishing user rate as $N_{SC} \rightarrow \infty$ with $N_{BS} = \lambda N_{SC}$, the SC and BS transmit power scaling factors χ_1 and η_1 must satisfy $0 \leq \chi_1 \leq 1$ and $0 \leq \eta_1 \leq 1$. When $\chi_1 = \eta_1 = 1$, the asymptotic achievable rate expressions of the i th ($i \in I$) MUE and j th ($j \in J_m$) SUE in the m th ($m \in \{1, \dots, S\}$)

$$\begin{aligned} \text{EEI}_i &= \alpha_{\text{BS}}^2 \left| \left(\xi_{\text{B-M}}^{(i)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right|^2, \quad \text{IMI}_i = \sum_{k \neq i}^K \alpha_{\text{BS}}^2 \left| \left(\xi_{\text{B-M}}^{(i)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(k)} \right|^2 \\ \text{CTI}_i &= \sum_{n=1}^S \sum_{l=1}^L \left(\alpha_{\text{SC}}^{(n)} \right)^2 \left| \left(\mathbf{g}_{\text{S-M}}^{(n,i)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(n,l)} \right|^2 \end{aligned} \quad (24)$$

$$\begin{aligned} \text{EEI}_{m,j} &= \left(\alpha_{\text{SC}}^{(m)} \right)^2 \left| \left(\xi_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right|^2, \quad \text{CTI}_{m,j} = \sum_{i=1}^K \alpha_{\text{BS}}^2 \left| \left(\mathbf{g}_{\text{B-S}}^{(m,j)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right|^2 \\ \text{ISI}_{m,j} &= \sum_{l_1 \neq j}^L \left(\alpha_{\text{SC}}^{(m)} \right)^2 \left| \left(\xi_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,l_1)} \right|^2, \quad \text{SSI}_{m,j} = \sum_{n \neq m}^S \sum_{l_2=1}^L \left(\alpha_{\text{SC}}^{(n)} \right)^2 \left| \left(\mathbf{g}_{\text{S-S}}^{(n,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(n,n,l_2)} \right|^2. \end{aligned} \quad (25)$$

$$R_{0,\text{M},\text{M}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{BS}} \left(\hat{\beta}_{\text{B-M}}^{(i)} \right)^2}{\Phi_{\text{B-M}} \sigma_0^2} \right), \quad R_{0,\text{S},\text{M}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{SC}} \left(\hat{\beta}_{\text{S-S}}^{(m,m,j)} \right)^2 / \Phi_{\text{S-S}}^{(m)}}{\sigma_0^2 + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \frac{E_{\text{SC}} \left(\hat{\beta}_{\text{S-S}}^{(n,n,j)} \right)^2 \beta_{\text{S-S}}^{(n,m,j)}}{\Phi_{\text{S-S}}^{(n)} \left(\beta_{\text{S-S}}^{(n,n,j)} \right)^2}} \right) \quad (28)$$

$$R_{0,\text{M},\text{Z}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{BS}}}{\Psi_{\text{B-M}} \sigma_0^2} \right), \quad R_{0,\text{S},\text{Z}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{SC}} / \Psi_{\text{S-S}}^{(m)}}{\sigma_0^2 + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \frac{E_{\text{SC}} \left(\beta_{\text{S-S}}^{(n,m,j)} \right)^2}{\Psi_{\text{S-S}}^{(n)} \left(\beta_{\text{S-S}}^{(n,n,j)} \right)^2}} \right) \quad (29)$$

$$R_{0,\text{M},\text{M}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{\lambda^\theta \tau E_\tau E_{\text{BS}} \left(\beta_{\text{B-M}}^{(i)} \right)^4}{\sum_{k=1}^K \left(\beta_{\text{B-M}}^{(k)} \right)^2 \sigma_0^4} \right), \quad R_{0,\text{S},\text{M}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{\tau E_\tau E_{\text{SC}} \left(\beta_{\text{S-S}}^{(m,m,j)} \right)^4}{\sum_{l=1}^L \left(\beta_{\text{S-S}}^{(m,m,l)} \right)^2 \sigma_0^4} \right) \quad (30)$$

$$R_{0,\text{M},\text{Z}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{\lambda^\theta \tau E_\tau E_{\text{BS}}}{\sum_{k=1}^K \left(\beta_{\text{B-M}}^{(k)} \right)^{-2} \sigma_0^4} \right), \quad R_{0,\text{S},\text{Z}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{\tau E_\tau E_{\text{SC}}}{\sum_{l=1}^L \left(\beta_{\text{S-S}}^{(m,m,l)} \right)^{-2} \sigma_0^4} \right) \quad (31)$$

SC for imperfect CSI based MRT and ZFT are (28) and (29), respectively, which show that the transmit powers at both BS and SCs can be scaled down by up to $\frac{1}{N_{\text{SC}}}$ to maintain a given rate in case I. When $0 \leq \chi_1 < 1$ and $0 \leq \eta_1 < 1$, the asymptotic achievable rate of each user approaches to infinity as $N_{\text{SC}} \rightarrow \infty$.

Remark 4: Obviously, when the pilot reuse factor $\gamma = S$, i.e., no pilot reuse, we have $R_{0,\text{S},\text{M}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{SC}} \left(\hat{\beta}_{\text{S-S}}^{(m,m,j)} \right)^2}{\Phi_{\text{S-S}}^{(m)} \sigma_0^2} \right)$ and

$R_{0,\text{S},\text{Z}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{SC}}}{\Psi_{\text{S-S}}^{(m)} \sigma_0^2} \right)$. For practical system configurations, we suppose that $\chi_1 \geq \eta_1$ in case I to guarantee $p_{\text{SC}} < p_{\text{BS}}$. When $0 < \chi_1 < 1$ and $0 < \eta_1 < 1$, the asymptotic achievable rate of both MUEs and SUEs approaches to infinity as $N_{\text{SC}} \rightarrow \infty$ for both MRT and ZFT. When $\chi_1 = 1$ and $\eta_1 = 0$, we have $R_{0,\text{M},\text{M}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \infty$, $R_{0,\text{M},\text{Z}}^{(i)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \infty$,

$$R_{0,\text{S},\text{M}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{\frac{E_{\text{SC}} \left(\hat{\beta}_{\text{S-S}}^{(m,m,j)} \right)^2}{\Phi_{\text{S-S}}^{(m)} E_{\text{BS}}}}{\beta_{\text{B-S}}^{(m,j)} + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \frac{E_{\text{SC}} \left(\hat{\beta}_{\text{S-S}}^{(n,n,j)} \right)^2 \beta_{\text{S-S}}^{(n,m,j)}}{E_{\text{BS}} \Phi_{\text{S-S}}^{(n)} \left(\beta_{\text{S-S}}^{(n,n,j)} \right)^2} + E_{\text{BS}}} \right),$$

$$\text{and } R_{0,\text{S},\text{Z}}^{(m,j)} \xrightarrow[N_{\text{SC}} \rightarrow \infty]{a.s.} \log_2 \left(1 + \frac{E_{\text{SC}} / \Psi_{\text{S-S}}^{(m)}}{\beta_{\text{B-S}}^{(m,j)} E_{\text{BS}} + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \frac{E_{\text{SC}} \left(\beta_{\text{S-S}}^{(n,m,j)} \right)^2}{\Psi_{\text{S-S}}^{(n)} \left(\beta_{\text{S-S}}^{(n,n,j)} \right)^2} + \sigma_0^2} \right),$$

respectively, indicating that the cross-tier interferences at SUEs can not be eliminated when p_{SC} is scaled down proportionally to $\frac{1}{N_{\text{SC}}}$ with fixed p_{BS} in case I.

Proposition 2: In case II where $p_\tau = \frac{E_\tau}{N_{\text{SC}}^\theta}$, $p_{\text{SC}}^{(s)} = p_{\text{SC}} = \frac{E_{\text{SC}}}{N_{\text{SC}}^{\chi_2}}$ ($s = 1, \dots, S$), $p_{\text{BS}} = \frac{E_{\text{BS}}}{N_{\text{BS}}^{\eta_2}}$, and E_τ , E_{BS} and E_{SC} are fixed, to achieve non-vanishing user rate as $N_{\text{SC}} \rightarrow \infty$ with $N_{\text{BS}} = \lambda N_{\text{SC}}$ and the pilot reuse factor $\gamma = S$, the pilot, SC and BS transmit power scaling factors θ , χ_2 and η_2 must

satisfy $0 < \theta \leq 1$, $0 \leq \chi_2 \leq 1 - \theta$ and $0 \leq \eta_2 \leq 1 - \theta$. When $0 < \theta < 1$ and $\chi_2 = \eta_2 = 1 - \theta$, the asymptotic achievable rate expressions of the i th ($i \in I$) MUE and j th ($j \in J_m$) SUE in the m th ($m \in \{1, \dots, S\}$) SC for imperfect CSI based MRT and ZFT are (30) and (31), respectively, from which we conclude that the transmit powers of BS and SCs can only be reduced by up to $\frac{1}{N_{SC}^{1-\theta}}$ with the pilot transmit power set as $p_\tau = \frac{E_\tau}{N_{SC}^\theta}$ and pilot reuse factor $\gamma = S$ (no pilot reuse) in case II. When $0 < \theta < 1$, $0 \leq \chi_2 < 1 - \theta$ and $0 \leq \eta_2 < 1 - \theta$, the asymptotic achievable rate of each user approaches to infinity as $N_{SC} \rightarrow \infty$.

Remark 5: To guarantee MUE and SUE achievable rate, $\chi_2 = \eta_2 = 0$ should be satisfied in case II when $\theta = 1$, which means that the pilot power can be scaled down by up to $\frac{1}{N_{SC}}$ with fixed transmit power at both BS and SC nodes. Then, the asymptotic achievable rate can be expressed as

$$\begin{aligned} R_{0,M,M}^{(i)} &\xrightarrow{a.s.} \log_2 \left(1 + \frac{\lambda \tau E_\tau E_{BS} (\beta_{B-M}^{(i)})^4}{\sum_{k=1}^K (\beta_{B-M}^{(k)})^2 \sigma_0^2 (RD_i + \sigma_0^2)} \right), \\ R_{0,S,M}^{(m,j)} &\xrightarrow{a.s.} \log_2 \left(1 + \frac{\tau E_\tau E_{SC} (\beta_{S-S}^{(m,j)})^4}{\sum_{l=1}^L (\beta_{S-S}^{(m,l)})^2 \sigma_0^2 (RD_{m,j} + \sigma_0^2)} \right), \\ R_{0,M,Z}^{(i)} &\xrightarrow{a.s.} \log_2 \left(1 + \frac{\lambda \tau E_\tau E_{BS} / \sum_{k=1}^K (\beta_{B-M}^{(k)})^{-2}}{\sigma_0^2 (RD_i + \sigma_0^2)} \right) \quad \text{and} \\ R_{0,S,Z}^{(m,j)} &\xrightarrow{a.s.} \log_2 \left(1 + \frac{\tau E_\tau E_{SC} / \sum_{l=1}^L (\beta_{S-S}^{(m,l)})^{-2}}{\sigma_0^2 (RD_{m,j} + \sigma_0^2)} \right) \quad \text{for} \end{aligned}$$

MRT and ZFT, respectively, where the residual items are $RD_i = E_{BS} \beta_{B-M}^{(i)} + E_{SC} \sum_{n=1}^S \beta_{S-M}^{(n,i)}$ and $RD_{m,j} = E_{SC} \sum_{n=1}^M \beta_{S-S}^{(n,m,l)} + E_{BS} \beta_{B-S}^{(m,j)}$. It indicates that the channel estimation error induced interference, cross-tier and inter-SC interferences can not be eliminated when p_τ is scaled down proportionally to $\frac{1}{N_{SC}}$ with fixed p_{BS} and p_{SC} in case II.

Proposition 3: In case II as stated in Proposition 2, to achieve non-vanishing user rate as $N_{SC} \rightarrow \infty$ with $N_{BS} = \lambda N_{SC}$ and the pilot reuse power $\gamma < S$, i.e., considering the pilot reuse introduced contamination, the pilot power scaling factor must satisfy $\theta = 0$. If $\theta > 0$, the MUE rate $R_{0,M,M}^{(i)}$ and $R_{0,M,Z}^{(i)}$ still follow (30) and (31), while the asymptotic achievable rate expressions of the j th ($j \in J_m$) SUE in the m th ($m \in \{1, \dots, S\}$) SC for imperfect CSI based MRT and ZFT are

$$R_{0,S,M}^{(m,j)} \xrightarrow[N_{SC} \rightarrow \infty]{a.s.} 0, \quad R_{0,S,Z}^{(m,j)} \xrightarrow[N_{SC} \rightarrow \infty]{a.s.} 0. \quad (32)$$

IV. USER SCHEDULING ALGORITHMS

To maximize the sum rate of the scheduled MUEs and SUEs, exhaustive search in the whole MUE and SUE sets is one possible method to obtain optimal results. However, it is not practical since it has rather low searching speed with high complexity. In this section, $\gamma = S$, i.e., no pilot reuse,

is assumed². As a traditional suboptimal method, a greedy scheduling algorithm is proposed according to the derived capacity lower bounds in Section III. Then, in comparison to the greedy scheduling algorithm, we propose a much simpler scheduling algorithm based on the obtained asymptotic results to maximize each cell's sum rate supposing that $N_{SC} \rightarrow \infty$. It is called asymptotic scheduling algorithm which significantly reduces the computation complexity.

A. Greedy Scheduling Algorithm (GSA)

Aiming to obtain an optimal user scheduling algorithm, we formulate an optimization problem considering the maximization of total achievable rate for all the scheduled MUEs and SUEs, subject to the constraints on the scale of each UE set, i.e., (33), which can surely be solved by inefficient exhaustive search. To reduce the computation complexity, a greedy scheduling algorithm is proposed, as summarized in Algorithm 1. In one iteration, each cell schedules one additional user, the user that maximizes the total sum rate is added in one cell, given the scheduled users in all other cells. This process repeats until the number of scheduled users in each cell reaches the target value.

As shown above, the proposed GSA is a suboptimal solution for the maximization of the sum rate, but it still results in high computation complexity, since each user's achievable rate is determined by the SCSIs of the downlink channels from the BS and all SC nodes to the users, i.e., the BS and each SC node should share SCSIs with one another even for each cell's own user scheduling.

Algorithm 1: Greedy scheduling algorithm

Initialization: $N = \lfloor \frac{K}{L} \rfloor$, $I = \emptyset$, $J_m = \emptyset$, $\tilde{U}_M = U_M$ and $\tilde{U}_S^{(m)} = U_S^{(m)}$ for $m = 1, \dots, S$.
Repeat:
 For1 $k = 1$ to N
 $i^* = \arg \max_{i \in \tilde{U}_M} R_{\text{SUM}}(I \cup \{i\}, J_1, \dots, J_S)$
 $I = I \cup \{i^*\}$, $\tilde{U}_M = \tilde{U}_M \setminus \{i^*\}$
 Endfor1
 For2 $m = 1$ to S
 $j_m^* = \arg \max_{j_m \in \tilde{U}_S^{(m)}} R_{\text{SUM}}(I, J_1, \dots, J_m \cup \{j_m\}, \dots, J_S)$
 $J_m = J_m \cup \{j_m^*\}$, $\tilde{U}_S^{(m)} = \tilde{U}_S^{(m)} \setminus \{j_m^*\}$
 Endfor2
Until: If $|J_m| = L$ with $m = 1, \dots, S$, change $N = K \bmod (L)$, go through For1 Loop and then stop;
Output: Output I and J_m ($m = 1, \dots, S$) as the solutions.

B. Asymptotic Scheduling Algorithm (ASA)

In order to further reduce the computation complexity of the traditional GSA, a new algorithm named ASA is presented for

²For other pilot reuse factors, the corresponding user scheduling algorithm can be a similar way.

$$\begin{aligned} \max_{I \subseteq U_M, J_m \subseteq U_S^{(m)}} R_{\text{SUM}}(I, J_1, \dots, J_S) &\triangleq \frac{T-\tau}{T} \left(\sum_{i \in I} R_{0,M}^{(i)} + \sum_{m=1}^S \sum_{j \in J_m} R_{0,S}^{(m,j)} \right) \\ \text{s.t. } |I| &= K, \quad |J_m| = L, \quad m = 1, \dots, S \end{aligned} \quad (33a)$$

$$(33b)$$

MRT and ZFT, respectively, based on the asymptotic results as given in Subsection III-C. From Propositions 1 and 2, it can be concluded that there is no inter-cell interference when $N_{SC} \rightarrow \infty$, i.e., each cell is able to do user scheduling according to its own statistical CSI and there is no information exchange requirement among the BS and SC nodes any more.

1) *Asymptotic Scheduling Algorithm for MRT (ASA-M)*: Since the Propositions 1 and 2 provide the asymptotic results for MRT, we define the achievable rate of the i th ($i \in U_{BS}$) MUE and j th ($j \in U_{SC}^{(m)}$) SUE in the m th ($m = 1, \dots, S$) SC as $R_{0,M,M-AS}^{(i)}$ and $R_{0,S,M-AS}^{(m,j)}$ according to (28) or (30) in Subsection III-C. Since maximizing the sum rate is equivalent to maximizing each cell's rate, optimization problems for the MC and each SC are proposed on the constraints of the sizes for each selected user subset, respectively, i.e.,

$$\max_{I \subseteq U_M} R_{BS,M}(I) \triangleq \frac{T-\tau}{T} \sum_{i \in I} R_{0,M,M-AS}^{(i)} \quad (34a)$$

$$\text{s.t. } |I| = K \quad (34b)$$

$$\max_{J_m \subseteq U_S^{(m)}} R_{SC,M}^{(m)}(J_m) \triangleq \frac{T-\tau}{T} \sum_{j \in J_m} R_{0,S,M-AS}^{(m,j)} \quad (35a)$$

$$\text{s.t. } |J_m| = L \quad (35b)$$

which can be solved separately by exhaustive search. Similarly, as a suboptimal solution, a simplified greedy search method is proposed, in which the user scheduling at each cell is operated separately, i.e., it can be completed by its own node without sharing any CSI with other cells. The reader is referred to Algorithm 2 for a step-by-step summary of the proposed method.

Algorithm 2: Asymptotic scheduling algorithm for MRT

Initialization: $I = \emptyset$, $J_m = \emptyset$, $\tilde{U}_M = U_M$ and $\tilde{U}_S^{(m)} = U_S^{(m)}$ for $m = 1, \dots, S$.

Repeat1:

$$i^* = \arg \max_{i \in \tilde{U}_M} R_{BS,M}(I \cup \{i\})$$

$$I = I \cup \{i^*\}, \quad \tilde{U}_M = \tilde{U}_M \setminus \{i^*\}$$

Until1: Stop if $|I| = K$.

For $m = 1$ to S

Repeat2:

$$j_m^* = \arg \max_{j_m \in \tilde{U}_S^{(m)}} R_{SC,M}^{(m)}(J_m \cup \{j_m\})$$

$$J_m = J_m \cup \{j_m^*\}, \quad \tilde{U}_S^{(m)} = \tilde{U}_S^{(m)} \setminus \{j_m^*\}$$

Until2: Stop if $|J_m| = L$.

Endfor

Output: Output I and J_m ($m = 1, \dots, S$) as the solutions.

2) *Asymptotic Scheduling Algorithm for ZFT (ASA-Z)*:

Likewise, since the Propositions 1 and 2 also provide the asymptotic results for ZFT, we can define the achievable rate of the i th ($i \in U_{BS}$) MUE and j th ($j \in U_{SC}^{(m)}$) SUE in the m th ($m = 1, \dots, S$) SC as $R_{0,M,Z-AS}^{(i)}$ and $R_{0,S,Z-AS}^{(m,j)}$ according to (29) or (31) in Subsection III-C. However, notably, $R_{0,M,Z-AS}^{(1)} = \dots = R_{0,M,Z-AS}^{(K)}$ for fixed $\sum_{k=1}^K (\beta_{B-M}^{(k)})^{-\eta}$, and $R_{0,S,Z-AS}^{(m,1)} = \dots = R_{0,S,Z-AS}^{(m,L)}$ for fixed $\sum_{l=1}^L (\beta_{S-S}^{(m,m,l)})^{-\eta}$ with $m = 1, \dots, S^3$, which means that maximizing the sum rate is equivalent to

$$\min_{I \subseteq U_M} \sum_{k \in I} (\beta_{B-M}^{(k)})^{-\eta} \quad (36a)$$

$$\text{s.t. } |I| = K \quad (36b)$$

$$\min_{J_m \subseteq U_S^{(m)}} \sum_{l \in J_m} (\beta_{S-S}^{(m,m,l)})^{-\eta} \quad (37a)$$

$$\text{s.t. } |J_m| = L. \quad (37b)$$

Different from ASA-M, the optimal ASA-Z can be realized by the max-beta based scheduling algorithm, the process of which is similar to the Algorithm 2. To solve this optimization problems of (36) and (37), we just need to find K MUEs with the first K -maximum $\beta_{B-M}^{(k)}$ and L SUEs with the first L -maximum $\beta_{S-S}^{(m,m,l)}$, respectively. Hence, we replace the item $R_{BS,M}(I \cup \{i\})$ and $R_{SC,M}^{(m)}(J_m \cup \{j_m\})$ in the Algorithm 2 by $\beta_{B-M}^{(i)}$ and $\beta_{S-S}^{(m,m,j_m)}$, respectively, to realize the optimal ASA-Z, which is much simpler than ASA-M as neither achievable rate calculation nor SCS exchange is needed at each node.

V. NUMERICAL RESULTS

In this section, we present numerical results to validate the derived achievable rate expressions and examine the performance of HetNet with large-scale antenna arrays in downlink channels. A dense tier of S SCs that are uniformly distributed in a circle as shown in Fig. 1. Suppose that the users in the cell are uniformly distributed with the total number 200 and 500 for $S = 8$ and $S = 20$, respectively, and that the MUEs and SUEs are identified by the bias user association. The number of scheduled MUEs and SUEs in each SC are $K = 20$ and $L = 4$, respectively, with training length $\tau = K + L \times \gamma$ and bias factors $\kappa_{BS} = 1$ and $\kappa_{SC} = 1.2$ [3] for BS and SC nodes, respectively. Also we assume that equal transmit power at the each SC node satisfies $p_{SC}^{(1)} \text{ (dBm)} = \dots = p_{SC}^{(S)} \text{ (dBm)} = p_{BS} \text{ (dBm)} - 22 \text{ (dB)}$ as p_{BS}

³Here, $\eta = 1$ for case I in (29) and $\eta = 2$ for case II in (31).

changes, and that $N_{BS}/N_{SC} = \lambda = 10$ is fixed as N_{BS} changes. See Table I for simulation parameters and assumption details.

TABLE I: SIMULATION PARAMETERS

Parameters	Setting
Bandwidth	20 MHz
Macro Cell radius	1000 m
Distance (BS,SCs)	800 m
The number of SCs	$S = 8$ or 20
Pilot Reuse factor	$\gamma = 1 \sim 8$ or $1 \sim 20$
Coherent interval	$T = 200$
Transmit power threshold (BS/SC)	$p_{BS} - p_{SC} = 22$ dB, $p_{SC}^{(1)} = \dots = p_{SC}^{(S)} = p_{SC}$
Noise power density	-174 dBm/Hz
Number of MUEs/SUEs	$K = 20$, $L = 4$
Pathloss model (BS)	$\theta_{BS}(d) = 128.1 + 37.6 \log_{10}(d)$, d (km) [26]
Pathloss model (SC)	$\theta_{SC}(d) = 140.7 + 36.7 \log_{10}(d)$, d (km) [26]

A. Comparison between One-Tier and Two-Tier Network Topologies

First, the effectiveness of the considered large-scale two-tier model is demonstrated by comparing it with the one-tier network topology. Here, suppose that simple random scheduling algorithm (RSA) is utilized for user scheduling, $p_{\tau} = 0$ dBm, $S = 8$ and $\gamma = 8$. To be fair, in the one-tier network the number of antennas is assumed to satisfy $N = N_{BS} + S \times N_{SC}$ and the number of UEs is set to be $K + S \times L$ including both MUEs and SUEs in the two-tier network. Fig. 2 shows that the two-tier network outperforms one-tier network on the spectral efficiency with both MRT and ZFT precoding schemes in the downlink channels. As an indication of coverage, Fig. 3 plots the cell boundary small cell user rate versus p_{BS} with fixed $N_{BS} = 80$, 160 and 320, where the cell boundary users are composed of all the SUEs identified by the biased user association [3]. For MRT the two-tier network has much higher cell boundary user rate, indicating better coverage than one-tier. However, for ZFT the two-tier network underperforms the one-tier for cell boundary SUEs, since neither the inter-SC interference nor the cross-tier interference can be cancelled in the non-cooperative two-tier HetNet systems. While in the one-tier network, the inter-user interferences could be totally eliminated by the ZFT precoding without inter-SC and cross-tier interferences. On the other hand, the ZFT precoding of the one-tier network is much more complex than that of the two-tier network due to the larger number of UEs ($K + S \times L$) served simultaneously by the BS.

B. Validation of Lower Capacity Bounds and Pilot Reuse Pattern

In this subsection, the effectiveness of the derived capacity lower bounds with imperfect CSI based MRT in (18) and (19)

and ZFT in (26) and (27) is evaluated by comparing them with the Monte-Carlo simulation results. Then, the performance of the presented simple PR pattern is evaluated. Besides, the asymptotic analyses with massive arrays for the two cases in Propositions 1 and 2 are examined in this subsection. Here, random scheduling algorithm (RSA) is utilized.

Firstly, $p_{\tau} = 0$ dBm, $S = 8$ and $\gamma = 8$ are set. In Fig. 4, the spectral efficiency versus p_{BS} curves with fixed $N_{BS} = 80$, 160 and 320 for capacity lower bounds are compared with those obtained from (12) and (13) by Monte-Carlo simulation. “Lower Bound (MRT)” indicates the capacity lower bound obtained by (18) and (19), and “Lower Bound (ZFT)” is calculated from (26) and (27). It can be observed from Fig. 4 that the relative performance gap between “Lower Bound (ZFT)” and “Monte-Carlo (ZFT)” is quite smaller than that between “Lower Bound (MRT)” and “Monte-Carlo (MRT)”, especially at lower transmit power with larger number of transmit antennas. Moreover, the spectral efficiency of ZFT increases much faster than that of MRT as p_{BS} increases, due to the fact that the effect of interference is much larger than that of the noise for higher SNR while ZFT is able to null multi-user interference signals [27]. In this way, the derived capacity lower bounds by Jensen’s inequality are proved to be accurate predictors of the system performance.

Secondly, setting $p_{BS} = 46$ dBm, $p_{\tau} = 0$ dBm and $S = 20$, the spectral efficiency of the presented PR pattern with different pilot reuse factor γ versus the number of transmit antennas at BS is illustrated in Fig. 5. When the spectral efficiency is calculated by (18) and (19) taking the influence of the pilot overhead τ into consideration, Fig. 5 indicates that PR factor $\gamma = 4$ yields the best performance for both MRT and ZFT, which achieves the optimal trade off between pilot overhead and pilot contamination from the reuse of the pilot sets. Notably, when the PR factor is reduced to $\gamma = 2$ and $\gamma = 1$, the spectral efficiency of ZFT is no longer larger than that of MRT as shown in Fig. 5, which indicates that pilot reuse introduced pilot contamination effect is more severe for ZFT than that for MRT. It can be explained by the fact that more pilot reuse increases the channel estimation error, and thus monotonically decreases the SINR of the SUEs in the ZFT case as stated by Remark 2.

Finally, the asymptotic analyses with massive arrays for the two cases in Propositions 1 and 2 are examined with the MUE and SUE large-scale fading factors fixed as $\beta_{B-M} = 1$, $\beta_{B-S} = 0.2$, $\beta_{S-S}^{(m,m)} = 5$, $\beta_{S-S}^{(m,n)} = 0.6$ ($m \neq n$) and $\beta_{S-M} = 0.6$ considering pathloss, noise variance $\sigma_0^2 = 1$ and the normalized pilot power $p_{\tau} = 0$ dB. Fig. 6 shows the required MC and SC transmit power p_{BS} and p_{SC} to achieve 1 bit/s/Hz per MUE and SUE, respectively, for MRT in case I. It is obvious from Fig. 6 that in case I where the pilot power p_{τ} is fixed, the required p_{BS} and p_{SC} are significantly reduced as N_{SC} increases, and that the required p_{BS} with $\gamma = 1$ is the lowest and the p_{SC} with $\gamma = 4$ is the lowest in comparison to other PR factors. Moreover, for $\gamma < 4$ in Fig. 6, it is evident that the lower the PR factor γ is, the higher p_{SC} is required to achieve 1 bit/s/Hz per user. The observation indicates that the pilot contamination effect existed at the SCs increases the required transmit power p_{SC} to achieve 1 bit/s/Hz per SUE.

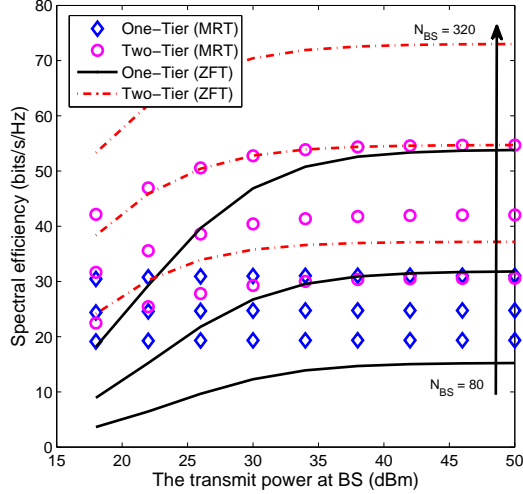


Fig. 2: Spectral efficiency versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

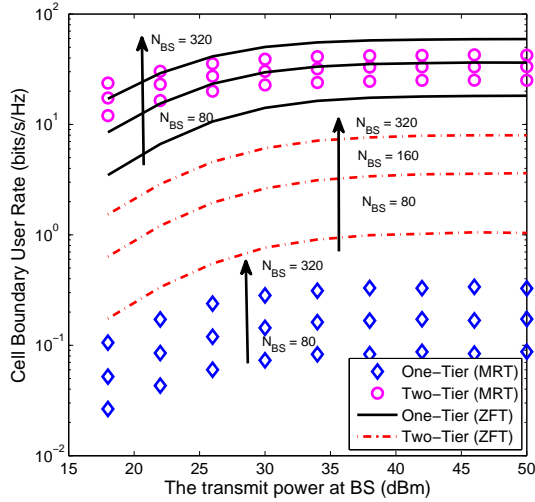


Fig. 3: Cell boundary SC user rate versus p_{BS} for one-tier and two-tier network topologies (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

Regarding the imperfect CSI effect, Fig. 7 shows that less transmit power in both MC and SCs are required when p_τ is high. For case II with $E_\tau = 0$ dB and the pilot power scaling down by $p_\tau = \frac{E_\tau}{N_{SC}^\tau}$, Fig. 8 shows that higher θ leads to more slowly reduced \bar{P}_τ^{SC} , because the imperfect CSI effect becomes more severe when the pilot power is reduced much faster with the increase of N_{SC} .

C. User Scheduling

In this subsection, the proposed user scheduling algorithms are examined with respect to the spectral efficiency of the HetNet downlink systems. Here, we choose $p_\tau = 5$ dBm, $S = 8$ and $\gamma = 8$ for comparison fairness.

First, simulations are performed for MRT on the proposed two user scheduling algorithms in regard to the spectral

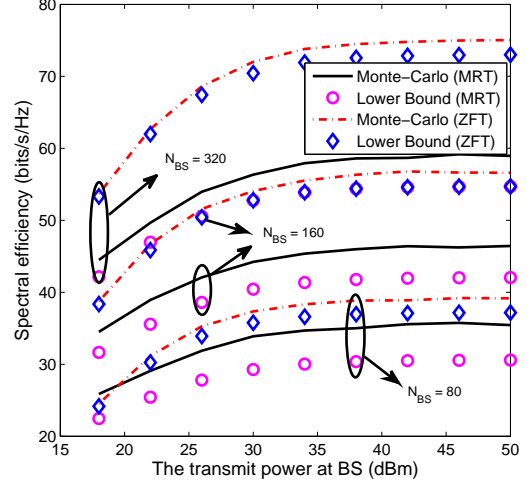


Fig. 4: Spectral efficiency versus p_{BS} for Monte-Carlo results and lower bounds (RSA, $p_\tau = 0$ dBm, $S = 8$, $\gamma = 8$).

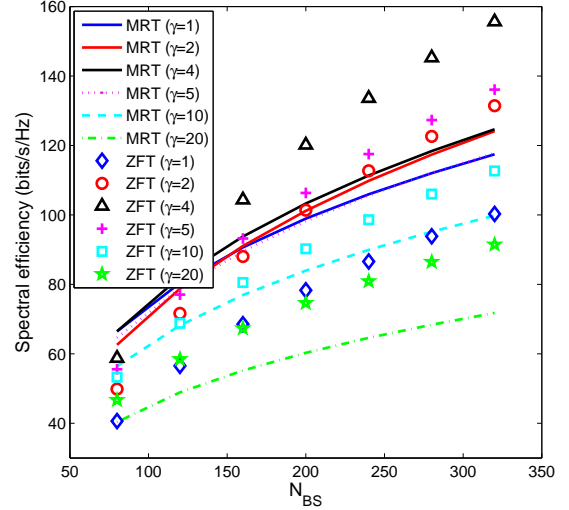


Fig. 5: Spectral efficiency versus N_{BS} for different PR factors (RSA, $p_{BS} = 46$ dBm, $p_\tau = 0$ dBm, $S = 20$).

efficiency versus p_{BS} , which is given in Fig. 9 with fixed $N_{BS} = 80, 160$ and 320 . We obtain the results that GSA outperforms ASA and RSA even when N_{BS} is not sufficiently high, and that the curve of ASA approaches to the curve of GSA as N_{BS} increases. Furthermore, the performance gap between ASA and RSA for MRT is large. Fig. 10 shows the spectral efficiency versus p_{BS} for ZFT on two different user scheduling schemes. It can be observed that the two curves for GSA and ASA almost coincide with each other even when the number of antennas is not very large, which demonstrates the effectiveness of ASA. Hence, for user scheduling of both MRT and ZFT, ASA is a good choice that achieves better performance with lower complexity. Moreover, by comparing Fig. 4 with Fig. 9 and Fig. 10, it is shown that increasing p_τ from 0 dBm to 5 dBm introduces better performance for ZFT than MRT, indicating that different imperfect CSI effect on

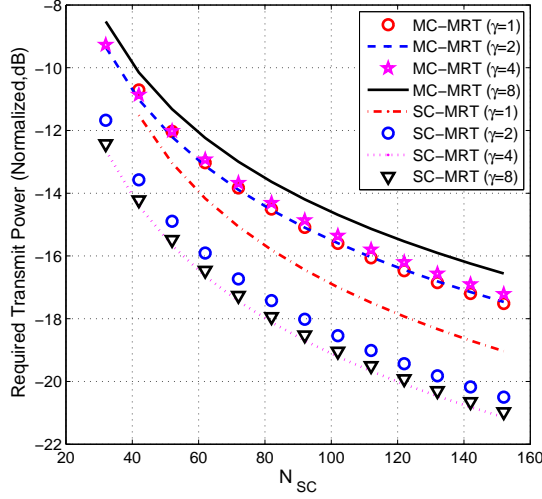


Fig. 6: Transmit power required to achieve 1 bit/s/Hz per user for case I (normalized $p_r = 0$ dB, $S = 8$).

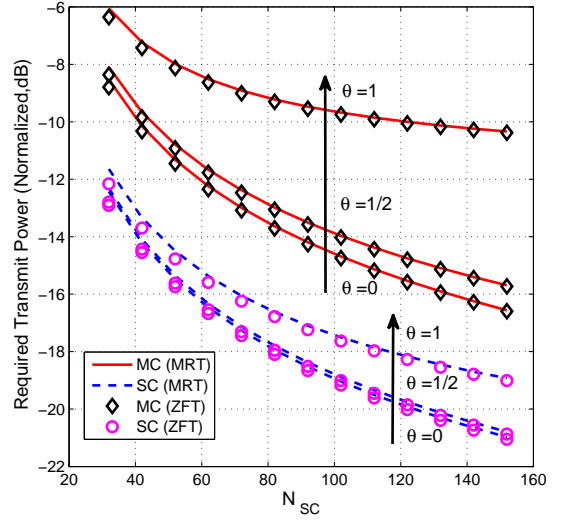


Fig. 8: Transmit power required to achieve 1 bit/s/Hz per user for case II (normalized $E_r = 0$ dB, $S = 8$, $\gamma = 8$).

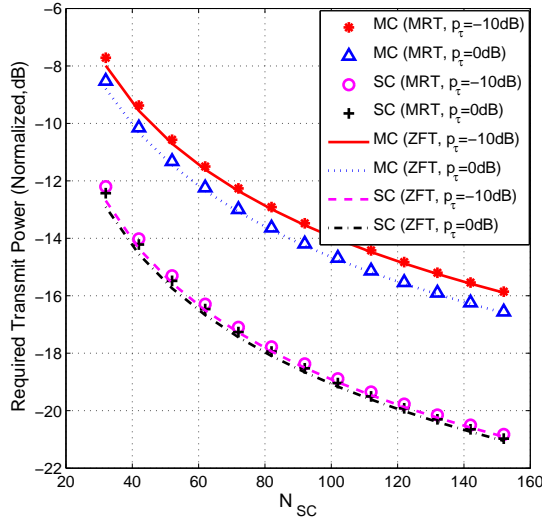


Fig. 7: Transmit power required to achieve 1 bit/s/Hz per user for case I ($S = 8$, $\gamma = 8$).

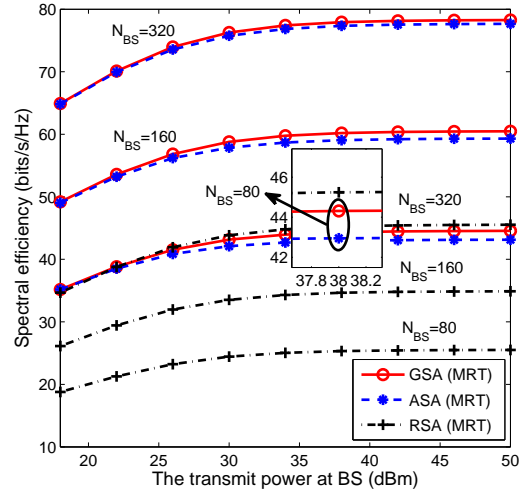


Fig. 9: Spectral efficiency versus p_{BS} for different user scheduling algorithms (MRT, $p_r = 5$ dBm, $S = 8$, $\gamma = 8$).

ZFT and MRT.

VI. CONCLUSION

In this paper, we have investigated the performance of a two-tier network with large-scale antenna arrays set at both BS and SCs. With MRT precoding employed at each node, we have derived capacity lower bounds with closed-form expressions for both imperfect CSI based MRT and ZFT cases, where a simple pilot reuse pattern is utilized for the channel estimation procedure to obtain the estimated imperfect CSI, followed by asymptotic analyses. The benefits of employing large number of antennas at both BS and SCs have been demonstrated. Simulation results have shown that the derived closed-form expressions for the achievable rate are accurate predictors of the system performance for both MRT and ZFT, and that

more pilot reuse is able to yield higher effective performance under different system configurations even though pilot contamination exists. As for user scheduling, two schemes have been proposed. The greedy scheduling algorithm designed based on the derived capacity lower bounds only requires the statistical CSI but not the instantaneous CSI exchange among BS and SCs. The asymptotic scheduling algorithm, based on the asymptotic analysis results, has even lower complexity by removing the need for any CSI exchange among nodes, and can still achieve near-optimal performance for both MRT and ZFT in the asymptotic regime of massive antenna arrays. Furthermore, it has been found that in the asymptotic regime and when ZFT is used, the capacity lower bound of a user is proportional to the large-scale fading factor of the channel from the user to its base station. Therefore, ASA-Z can further reduce complexity by circumventing the need to calculate the

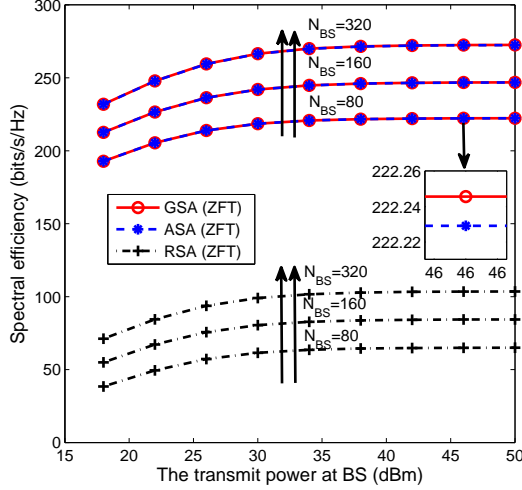


Fig. 10: Spectral efficiency versus p_{BS} for different user scheduling algorithms (ZFT, $p_\tau = 5$ dBm, $S = 8$, $\gamma = 8$).

achievable rate.

VII. PROOF OF (9)

As to the derivation of (9), we use the identity [30, 31]

$$\mathbb{E}[\mathbf{W}^{-1}] = \frac{\mathbf{\Sigma}^{-1}}{n - m - 1} \quad (38)$$

where $\mathbf{W} \sim \mathcal{W}_m(\mathbf{\Sigma}, n)$ is an $m \times m$ central complex Wishart matrix with n ($n > m$) degrees of freedom and the distribution of \mathbf{W}^{-1} is called an inverted Wishart distribution, following $\mathcal{W}_m^{-1}(\mathbf{\Sigma}^{-1}, n)$. It can be easily concluded that $(\hat{\mathbf{G}}_{B-M}^H \hat{\mathbf{G}}_{B-M})^{-1} \sim \mathcal{W}_K^{-1}(\hat{\mathbf{D}}_{B-M}^{-1}, N_{BS})$ and $\left[(\hat{\mathbf{G}}_{S-S}^{(m,m)})^H \hat{\mathbf{G}}_{S-S}^{(m,m)} \right]^{-1} \sim \mathcal{W}_L^{-1}(\hat{\mathbf{D}}_{S-S}^{(m,m)})^{-1}, N_{SC})$ for $\forall m \in \{1, \dots, S\}$, hence

$$\begin{aligned} \mathbb{E} \left[(\hat{\mathbf{G}}_{B-M}^H \hat{\mathbf{G}}_{B-M})^{-1} \right] &= \frac{\hat{\mathbf{D}}_{B-M}^{-1}}{N_{BS} - K - 1} \\ \mathbb{E} \left[\left[(\hat{\mathbf{G}}_{S-S}^{(m,m)})^H \hat{\mathbf{G}}_{S-S}^{(m,m)} \right]^{-1} \right] &= \frac{(\hat{\mathbf{D}}_{S-S}^{(m,m)})^{-1}}{N_{SC} - L - 1} \end{aligned} \quad (39)$$

where $N_{BS} > K$ and $N_{SC} > L$. In this way, we have

$$\alpha_{BS} = \sqrt{\frac{(N_{BS} - K - 1) P_{BS}}{\Psi_{B-M}}}, \alpha_{SC}^{(m)} = \sqrt{\frac{(N_{SC} - L - 1) P_{SC}^{(m)}}{\Psi_{S-S}^{(m)}}} \quad (40)$$

where $\Psi_{B-M} = \sum_{i=1}^K \frac{1}{\hat{\beta}_{B-M}^{(i)}}$, and $\Psi_{S-S}^{(m)} = \sum_{l=1}^L \frac{1}{\hat{\beta}_{S-S}^{(m,l)}}$ with $m \in \{1, \dots, S\}$.

VIII. PROOF OF THEOREM 1

To derive the closed-form expression of the MUE achievable rate in (18) based on Jensen's inequality, we start from the expectation of the SINR's reciprocal, given by (41), where

the item of $\frac{\mathbb{E}[\text{CTI}_i] + \sigma_0^2}{\alpha_{BS}^2}$ does not depend on $\hat{\mathbf{g}}_{B-M}^{(i)}$. Since large-scale antenna arrays are set at both the BS and SCs, some results from Gaussian distributed estimated channel in (3)⁴ [28] can be utilized. Using [29, Lemma 2.9] (42), where $\mathbf{W} \sim \mathcal{W}_m(\mathbf{I}_n, n)$ is an $m \times m$ central complex Wishart matrix with n ($n > m + 1$) degrees of freedom, we have (43).

Then, we define $\tilde{\xi}_{B-M}^{(i)} \triangleq \frac{(\xi_{B-M}^{(i)})^H \hat{\mathbf{g}}_{B-M}^{(i)}}{\|\hat{\mathbf{g}}_{B-M}^{(i)}\|^2} \sim \mathcal{CN}(0, \beta_{B-M}^{(i)} - \hat{\beta}_{B-M}^{(i)})$ and $\hat{g}_{B-M}^{(k)} \triangleq \frac{(\hat{\mathbf{g}}_{B-M}^{(i)})^H \hat{\mathbf{g}}_{B-M}^{(k)}}{\|\hat{\mathbf{g}}_{B-M}^{(i)}\|^2} \sim \mathcal{CN}(0, \hat{\beta}_{B-M}^{(k)})$. Conditioned on $\hat{\mathbf{g}}_{B-M}^{(i)}$, $\tilde{\xi}_{B-M}^{(i)}$ and $\hat{g}_{B-M}^{(k)}$ are Gaussian distributed and independent of $\hat{\mathbf{g}}_{B-M}^{(i)}$. Hence, we have (44) and (45). Also, law of large numbers [28] leads to

$$\mathbb{E}[\text{CTI}_i] = \sum_{n=1}^S \left(\alpha_{SC}^{(n)} \right)^2 N_{SC} \Phi_{S-S}^{(n)} \beta_{S-M}^{(n,i)}. \quad (46)$$

By substituting (44), (45) and (46) into (41), (16) leads to

$$\begin{aligned} R_{0,M,M}^{(i)} &= \\ \log_2 \left(1 + \frac{\alpha_{BS}^2 (N_{BS} - 1)(N_{BS} - 2) (\hat{\beta}_{B-M}^{(i)})^2}{\alpha_{BS}^2 \hat{\chi}_{BS}^{(i)} + \sum_{n=1}^S \left(\alpha_{SC}^{(n)} \right)^2 N_{SC} \Phi_{S-S}^{(n)} \beta_{S-M}^{(n,i)} + \sigma_0^2} \right) \end{aligned} \quad (47)$$

where $\hat{\chi}_{BS}^{(i)} = \Phi_{B-M} \left(N_{BS} \beta_{B-M}^{(i)} - 2\hat{\beta}_{B-M}^{(i)} \right) - (N_{BS} - 4) \left(\hat{\beta}_{B-M}^{(i)} \right)^2 - 2\hat{\beta}_{B-M}^{(i)} \beta_{B-M}^{(i)}$.

For the derivation of the closed-form expression for the SUE achievable rate in (19), the expectation of the SINR's reciprocal can be written as (48), where $\frac{\mathbb{E}[\text{CTI}_{m,j}] + \mathbb{E}[\text{SSI}_{m,j}] + \sigma_0^2}{(\alpha_{SC}^{(m)})^2}$ is independent of $\hat{\mathbf{g}}_{S-S}^{(m,m,j)}$. According to the [29, Lemma 2.9] given in (42), we have (49). Likewise, we define $\tilde{\xi}_{S-S}^{(m,m,j)} \triangleq \frac{(\xi_{S-S}^{(m,m,j)})^H \hat{\mathbf{g}}_{S-S}^{(m,m,j)}}{\|\hat{\mathbf{g}}_{S-S}^{(m,m,j)}\|^2} \sim \mathcal{CN}(0, \beta_{S-S}^{(m,m,j)} - \hat{\beta}_{S-S}^{(m,m,j)})$ and $\hat{g}_{S-S}^{(m,m,l_1)} \triangleq \frac{(\hat{\mathbf{g}}_{S-S}^{(m,m,j)})^H \hat{\mathbf{g}}_{S-S}^{(m,m,l_1)}}{\|\hat{\mathbf{g}}_{S-S}^{(m,m,j)}\|^2} \sim \mathcal{CN}(0, \hat{\beta}_{S-S}^{(m,m,l_1)})$, both of which do not depend on $\hat{\mathbf{g}}_{S-S}^{(m,m,j)}$ conditioned on it, then we obtain (50) and (51). Similarly, the law of large numbers [28] leads to (52), where $\hat{\beta}_{S-S}^{(n,m,j)} \triangleq \frac{\tau p_\tau (\beta_{S-S}^{(n,m,j)})^2}{\tau p_\tau \sum_{l \in \mathcal{A}_r} \beta_{S-S}^{(n,l,j)} + \sigma_0^2}$. Then,

substituting (50), (51), and (52) into (49), (17) leads to (53), where $\hat{\chi}_{SC}^{(m,j)} = \Phi_{S-S}^{(m)} \left(N_{SC} \beta_{S-S}^{(m,m,j)} - 2\hat{\beta}_{S-S}^{(m,m,j)} \right) - (N_{SC} - 4) \left(\hat{\beta}_{S-S}^{(m,m,j)} \right)^2 - 2\hat{\beta}_{S-S}^{(m,m,j)} \beta_{S-S}^{(m,m,j)}$, $\hat{\chi}_{BS}^{(m,j)} = N_{BS} \Phi_{B-M} \beta_{B-M}^{(m,j)}$, $\hat{\chi}_{SC,1}^{(n,m,j)} = N_{SC} \Phi_{S-S}^{(n)} \beta_{S-S}^{(n,m,j)}$, and $\hat{\chi}_{SC,2}^{(n,m,j)} = N_{SC}^2 \beta_{S-S}^{(n,m,j)} \hat{\beta}_{S-S}^{(n,m,j)}$.

Finally, by substituting (7) into (47) and (53), (18) and (19)

⁴Due to the estimated channel model in (3), we have that $\hat{\mathbf{g}}_i$ and $\hat{\mathbf{g}}_j$ are mutually independent $N \times 1$ vectors with $\forall i \neq j$ whose elements are i.i.d. zero-mean Gaussian distributed with variances $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$, respectively. Then, it can be concluded that $\mathbb{E}[\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_i] = N \hat{\sigma}_i^2$, $\mathbb{E}[\hat{\mathbf{g}}_j^H \hat{\mathbf{g}}_j] = N \hat{\sigma}_j^2$, and $\mathbb{E}[\hat{\mathbf{g}}_i^H \hat{\mathbf{g}}_j] = 0$. Also, we can obtain that $\mathbb{E}[\|\hat{\mathbf{g}}_i\|^2 \hat{\sigma}_i^2] = N \hat{\sigma}_i^4$.

$$\mathbb{E} \left[\frac{\text{EEI}_i + \text{IMI}_i + \text{CTI}_i + \sigma_0^2}{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] = \mathbb{E} \left[\frac{\text{EEI}_i}{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] + \mathbb{E} \left[\frac{\text{IMI}_i}{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] + \mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] \frac{\mathbb{E} [\text{CTI}_i] + \sigma_0^2}{\alpha_{\text{BS}}^2} \quad (41)$$

$$\mathbb{E} [\text{Tr} \{ \mathbf{W}^{-1} \}] = \frac{m}{n-m}, \quad \mathbb{E} [Tr^2 \{ \mathbf{W}^{-1} \}] = \frac{m}{n-m} \left(\frac{n}{(n-m)^2 - 1} + \frac{m-1}{n-m+1} \right) \quad (42)$$

$$\mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^2} \right] = \frac{1}{(N_{\text{BS}} - 1) \hat{\beta}_{\text{B-M}}^{(i)}}, \quad \mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] = \frac{1}{(N_{\text{BS}} - 1)(N_{\text{BS}} - 2) (\hat{\beta}_{\text{B-M}}^{(i)})^2}. \quad (43)$$

$$\mathbb{E} \left[\frac{\text{EEI}_i}{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] = \mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^2} \right] \mathbb{E} [\xi_{\text{B-M}}^{(i)2}] = \frac{\beta_{\text{B-M}}^{(i)} - \hat{\beta}_{\text{B-M}}^{(i)}}{(N_{\text{BS}} - 1) \hat{\beta}_{\text{B-M}}^{(i)}} \quad (44)$$

$$\begin{aligned} \mathbb{E} \left[\frac{\text{IMI}_i}{\alpha_{\text{BS}}^2 \|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] &= \mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^4} \right] \sum_{k \neq i}^K \mathbb{E} \left[\left| (\xi_{\text{B-M}}^{(i)})^T (\hat{\mathbf{g}}_{\text{B-M}}^{(k)})^* \right|^2 \right] \\ &+ \mathbb{E} \left[\frac{1}{\|\hat{\mathbf{g}}_{\text{B-M}}^{(i)}\|^2} \right] \sum_{k \neq i}^K \mathbb{E} [\hat{g}_{\text{B-M}}^{(k)2}] = \frac{\sum_{k \neq i}^K \hat{\beta}_{\text{B-M}}^{(k)}}{(N_{\text{BS}} - 1) \hat{\beta}_{\text{B-M}}^{(i)}} + \frac{N_{\text{BS}} \sum_{k \neq i}^K (\beta_{\text{B-M}}^{(i)} - \hat{\beta}_{\text{B-M}}^{(i)}) \hat{\beta}_{\text{B-M}}^{(k)}}{(N_{\text{BS}} - 1)(N_{\text{BS}} - 2) (\hat{\beta}_{\text{B-M}}^{(i)})^2}. \end{aligned} \quad (45)$$

are obtained and thus Theorem 1 is demonstrated.

IX. PROOF OF THEOREM 2

Similar to the proof of Theorem 1 in Appendix VIII, to derive the closed-form expression of the ZFT based MUE achievable rate in (26) based on Jensen's inequality, we start from the expectation of the SINR's reciprocal, given by (54).

Then, the property of the inverted Wishart Distribution in (39), the law of large numbers [28] and $\text{Tr} \{ \mathbf{A} \mathbf{B} \} = \text{Tr} \{ \mathbf{B} \mathbf{A} \}$ offer us (55) and (56), where $\Psi_{\text{B-M}} = \sum_{i=1}^K \frac{1}{\hat{\beta}_{\text{B-M}}^{(i)}}$, and $\Psi_{\text{S-S}}^{(n)} = \sum_{l=1}^L \frac{1}{\hat{\beta}_{\text{S-S}}^{(n,l)}}$ with $n \in \{1, \dots, S\}$.

By substituting (55) and (56) into (54), (16) leads to (57).

For the derivation of the closed-form expression for the SUE achievable rate in (27), the expectation of the SINR's reciprocal can be written as

$$\begin{aligned} &\mathbb{E} \left[\frac{\text{EEI}_{m,j} + \text{CTI}_{m,j} + \text{ISI}_{m,j} + \text{SSI}_{m,j} + \sigma_0^2}{(\alpha_{\text{SC}}^{(m)})^2} \right] \\ &= \frac{\mathbb{E} [\text{EEI}_{m,j} + \text{ISI}_{m,j}] + \mathbb{E} [\text{CTI}_{m,j}] + \mathbb{E} [\text{SSI}_{m,j}] + \sigma_0^2}{(\alpha_{\text{SC}}^{(m)})^2}. \end{aligned} \quad (58)$$

Similarly, according to the property of the inverted Wishart Distribution in (39), we have (59), (60) and (61).

Then, substituting (59), (60) and (61) into (58), (17) leads to (62), where $\hat{\eta}_{\text{SC}}^{(m,j)} = \frac{(\beta_{\text{S-S}}^{(n,m,j)} - \hat{\beta}_{\text{S-S}}^{(m,m,j)}) \Psi_{\text{S-S}}^{(m)}}{N_{\text{SC}} - L - 1}$, $\hat{\eta}_{\text{BS}}^{(m,j)} = \frac{\beta_{\text{B-S}}^{(m,j)} \Psi_{\text{B-M}}}{N_{\text{BS}} - K - 1}$, $\hat{\eta}_{\text{SC},1}^{(n,m,j)} = \frac{\beta_{\text{S-S}}^{(n,m,j)} \Psi_{\text{S-S}}^{(n)}}{N_{\text{SC}} - L - 1}$, and $\hat{\eta}_{\text{SC},2}^{(n,m,j)} =$

$\left[\left(\frac{\beta_{\text{S-S}}^{(n,m,j)}}{\hat{\beta}_{\text{S-S}}^{(n,n,j)}} \right)^2 - \frac{\beta_{\text{S-S}}^{(n,m,j)}}{(N_{\text{SC}} - L - 1) \hat{\beta}_{\text{S-S}}^{(n,n,j)}} \right]$. Finally, by substituting (9) into (57) and (62), (26) and (27) are obtained and thus Theorem 2 is demonstrated.

REFERENCES

- [1] Y. Dai and X. Dong, "Downlink performance and user scheduling of HetNet with large-scale antenna arrays," In *Proc. IEEE Global Commun. Conf. (GLOBECOM'15)*, Dec. 2015.
- [2] 3GPP, TR 36.814, v9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA), Further Advancements for E-UTRA Physical Layer Aspects," Mar. 2010.
- [3] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: a utility perspective," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1025–1039, June 2015.
- [4] S. Brueck, "Heterogeneous networks in LTE-Advanced," In *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, pp. 171–175, Nov. 2011.
- [5] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10–21, June 2011.
- [6] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *arXiv preprint arXiv:1405.2957*, 2014.
- [7] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [8] J. Zhu and H. Yang, "Interference control with beamforming coordination for two-tier femtocell networks and its performance analysis," In *Proc. IEEE Int. Conf. Commun. (ICC'11)*, June 2011.
- [9] Y. Dai, S. Jin, L. Pan, X. Gao, L. Jiang, and M. Lei, "Interference control based on beamforming coordination for heterogeneous network with RRH deployment," *IEEE Syst. J.*, vol. 99, no. 99, pp. 1–7, Apr. 2013.
- [10] Z. Xu, C. Yang, G. Y. Li, Y. Liu, and S. Xu, "Energy-efficient CoMP precoding in heterogeneous networks," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 1005–1017, Feb. 2014.
- [11] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

$$\begin{aligned} \mathbb{E} \left[\frac{\text{EEI}_{m,j} + \text{CTI}_{m,j} + \text{ISI}_{m,j} + \text{SSI}_{m,j} + \sigma_0^2}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] &= \mathbb{E} \left[\frac{\text{EEI}_{m,j}}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] \\ &+ \mathbb{E} \left[\frac{\text{ISI}_{m,j}}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] + \mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] \frac{\mathbb{E} [\text{CTI}_{m,j}] + \mathbb{E} [\text{SSI}_{m,j}] + \sigma_0^2}{\left(\alpha_{\text{SC}}^{(m)}\right)^2} \end{aligned} \quad (48)$$

$$\mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^2} \right] = \frac{1}{(N_{\text{SC}} - 1) \hat{\beta}_{\text{S-S}}^{(m,m,j)}}, \mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] = \frac{1}{(N_{\text{SC}} - 1) (N_{\text{SC}} - 2) \left(\hat{\beta}_{\text{S-S}}^{(m,m,j)} \right)^2}. \quad (49)$$

$$\mathbb{E} \left[\frac{\text{EEI}_{m,j}}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] = \mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^2} \right] \mathbb{E} \left[\left| \tilde{\xi}_{\text{S-S}}^{(m,m,j)} \right|^2 \right] = \frac{\beta_{\text{S-S}}^{(m,m,j)} - \hat{\beta}_{\text{S-S}}^{(m,m,j)}}{(N_{\text{SC}} - 1) \hat{\beta}_{\text{S-S}}^{(m,m,j)}} \quad (50)$$

$$\begin{aligned} \mathbb{E} \left[\frac{\text{ISI}_{m,j}}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] &= \mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^4} \right] \sum_{l_1 \neq j}^L \mathbb{E} \left[\left| \left(\xi_{\text{S-S}}^{(m,m,j)} \right)^T \left(\hat{\mathbf{g}}_{\text{S-S}}^{(m,m,l_1)} \right)^* \right|^2 \right] \\ &+ \mathbb{E} \left[\frac{1}{\left\| \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right\|^2} \right] \sum_{l_1 \neq j}^L \mathbb{E} \left[\left| \hat{g}_{\text{S-S}}^{(m,m,l_1)} \right|^2 \right] = \frac{\sum_{l_1 \neq j}^L \hat{\beta}_{\text{S-S}}^{(m,m,l_1)}}{(N_{\text{SC}} - 1) \hat{\beta}_{\text{S-S}}^{(m,m,j)}} + \frac{N_{\text{SC}} \sum_{l_1 \neq j}^L \left(\beta_{\text{S-S}}^{(m,m,j)} - \hat{\beta}_{\text{S-S}}^{(m,m,j)} \right) \hat{\beta}_{\text{S-S}}^{(m,m,l_1)}}{(N_{\text{SC}} - 1) (N_{\text{SC}} - 2) \left(\hat{\beta}_{\text{S-S}}^{(m,m,j)} \right)^2} \end{aligned} \quad (51)$$

$$\mathbb{E} [\text{CTI}_{m,j}] = \alpha_{\text{BS}}^2 N_{\text{BS}} \Phi_{\text{B-M}} \beta_{\text{B-S}}^{(m,j)}, \mathbb{E} [\text{SSI}_{m,j}] = \sum_{n \neq m}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 N_{\text{SC}} \Phi_{\text{S-S}}^{(n)} \beta_{\text{S-S}}^{(n,m,j)} + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \left(\alpha_{\text{SC}}^{(n)} \right)^2 N_{\text{SC}}^2 \beta_{\text{S-S}}^{(n,m,j)} \hat{\beta}_{\text{S-S}}^{(n,m,j)} \quad (52)$$

$$R_{0,\text{S},\text{M}}^{(m,j)} = \log_2 \left(1 + \frac{\left(\alpha_{\text{SC}}^{(m)}\right)^2 (N_{\text{SC}} - 1) (N_{\text{SC}} - 2) \left(\beta_{\text{S-S}}^{(m,m,j)}\right)^2}{\left(\alpha_{\text{SC}}^{(m)}\right)^2 \hat{\chi}_{\text{SC}}^{(m,j)} + \alpha_{\text{BS}}^2 \hat{\chi}_{\text{BS}}^{(m,j)} + \sum_{n \neq m}^S \left(\alpha_{\text{SC}}^{(n)}\right)^2 \hat{\chi}_{\text{SC},1}^{(n,m,j)} + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \left(\alpha_{\text{SC}}^{(n)}\right)^2 \hat{\chi}_{\text{SC},2}^{(n,m,j)} + \sigma_0^2} \right) \quad (53)$$

$$\mathbb{E} \left[\frac{\text{EEI}_i + \text{IMI}_i + \text{CTI}_i + \sigma_0^2}{\alpha_{\text{BS}}^2} \right] = \frac{\mathbb{E} [\text{EEI}_i + \text{IMI}_i] + \mathbb{E} [\text{CTI}_i] + \sigma_0^2}{\alpha_{\text{BS}}^2}. \quad (54)$$

$$\mathbb{E} [\text{EEI}_i + \text{IMI}_i] = \alpha_{\text{BS}}^2 \mathbb{E} \left[\left(\xi_{\text{B-M}}^{(i)} \right)^H \hat{\mathbf{G}}_{\text{B-M}} \hat{\mathbf{G}}_{\text{B-M}}^H \xi_{\text{B-M}}^{(i)} \right] = \alpha_{\text{BS}}^2 \left(\beta_{\text{B-M}}^{(i)} - \hat{\beta}_{\text{B-M}}^{(i)} \right) \frac{\Psi_{\text{B-M}}}{N_{\text{BS}} - K - 1} \quad (55)$$

$$\mathbb{E} [\text{CTI}_i] = \sum_{n=1}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 \beta_{\text{S-S}}^{(n,i)} \text{Tr} \left\{ \mathbb{E} \left[\left[\left(\hat{\mathbf{G}}_{\text{S-S}}^{(n,n)} \right)^H \hat{\mathbf{G}}_{\text{S-S}}^{(n,n)} \right]^{-1} \right] \right\} = \sum_{n=1}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 \beta_{\text{S-S}}^{(n,i)} \frac{\Psi_{\text{S-S}}^{(n)}}{N_{\text{SC}} - L - 1} \quad (56)$$

$$R_{0,\text{M},\text{Z}}^{(i)} = \log_2 \left(1 + \frac{\alpha_{\text{BS}}^2}{\alpha_{\text{BS}}^2 \frac{\left(\beta_{\text{B-M}}^{(i)} - \hat{\beta}_{\text{B-M}}^{(i)}\right) \Psi_{\text{B-M}}}{N_{\text{BS}} - K - 1} + \sum_{n=1}^S \left(\alpha_{\text{SC}}^{(n)}\right)^2 \beta_{\text{S-S}}^{(n,i)} \frac{\Psi_{\text{S-S}}^{(n)}}{N_{\text{SC}} - L - 1} + \sigma_0^2} \right). \quad (57)$$

$$\begin{aligned} \mathbb{E}[\text{EEI}_{m,j} + \text{ISI}_{m,j}] &= \mathbb{E} \left[\left(\alpha_{\text{SC}}^{(m)} \right)^2 \left| \left(\xi_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,j)} \right|^2 + \sum_{l_1 \neq j}^L \left(\alpha_{\text{SC}}^{(m)} \right)^2 \left| \left(\xi_{\text{S-S}}^{(m,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(m,m,l_1)} \right|^2 \right] \\ &= \frac{\left(\alpha_{\text{SC}}^{(m)} \right)^2 \left(\beta_{\text{S-S}}^{(m,m,j)} - \hat{\beta}_{\text{S-S}}^{(m,m,j)} \right) \Psi_{\text{S-S}}^{(m)}}{N_{\text{SC}} - L - 1} \end{aligned} \quad (59)$$

$$\mathbb{E}[\text{CTI}_{m,j}] = \mathbb{E} \left[\sum_{i=1}^K \alpha_{\text{BS}}^2 \left| \left(\mathbf{g}_{\text{B-S}}^{(m,j)} \right)^H \hat{\mathbf{g}}_{\text{B-M}}^{(i)} \right|^2 \right] = \frac{\alpha_{\text{BS}}^2 \beta_{\text{B-S}}^{(m,j)} \Psi_{\text{B-M}}}{N_{\text{BS}} - K - 1} \quad (60)$$

$$\begin{aligned} \mathbb{E}[\text{SSI}_{m,j}] &= \sum_{n \neq m}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 \mathbb{E} \left[\sum_{l_2=1}^L \left| \left(\mathbf{g}_{\text{S-S}}^{(n,m,j)} \right)^H \hat{\mathbf{g}}_{\text{S-S}}^{(n,n,l_2)} \right|^2 \right] = \frac{\sum_{n \neq m}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 \beta_{\text{S-S}}^{(n,m,j)} \Psi_{\text{S-S}}^{(n)}}{N_{\text{SC}} - L - 1} \\ &+ \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \left(\alpha_{\text{SC}}^{(n)} \right)^2 \left[\left(\frac{\beta_{\text{S-S}}^{(n,m,j)}}{\beta_{\text{S-S}}^{(n,n,j)}} \right)^2 - \frac{\beta_{\text{S-S}}^{(n,m,j)}}{(N_{\text{SC}} - L - 1) \hat{\beta}_{\text{S-S}}^{(n,n,j)}} \right]. \end{aligned} \quad (61)$$

$$R_{0,\text{S},\text{Z}}^{(m,j)} = \log_2 \left(1 + \frac{\left(\alpha_{\text{SC}}^{(m)} \right)^2}{\left(\alpha_{\text{SC}}^{(m)} \right)^2 \hat{\eta}_{\text{SC}}^{(m,j)} + \alpha_{\text{BS}}^2 \hat{\eta}_{\text{BS}}^{(m,j)} + \sum_{n \neq m}^S \left(\alpha_{\text{SC}}^{(n)} \right)^2 \hat{\eta}_{\text{SC},1}^{(n,m,j)} + \sum_{\substack{n \neq m \\ n \in \mathcal{A}_r}} \left(\alpha_{\text{SC}}^{(n)} \right)^2 \hat{\eta}_{\text{SC},2}^{(n,m,j)} + \sigma_0^2} \right) \quad (62)$$

- [12] F. Fernandes, A. Ashikhmin and T. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.
- [13] H. Q. Ngo, E. G. Larsson and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [14] A. Adhikary, E. A. Safadi, and G. Caire, "Massive MIMO and inter-tier interference coordination," In *Proc. Inf. Theory and its Applications (ITA'14)*, Feb. 2014.
- [15] K. Hosseini, J. Hoydis, S. Brink and M. Debbah, "Massive MIMO and small cells: how to densify heterogeneous networks," In *Proc. IEEE Int. Conf. Commun. (ICC'13)*, pp. 5442–5447, 2013.
- [16] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," In *Proc. IEEE Veh. Tech. Conf. (VTC'11)*, pp. 1–5, Sept. 2011.
- [17] "Massive-element antenna for small cell solutions in 5G," *White Paper* by NEC Cooperation, Feb. 2016.
- [18] A. Adhikary, H. S. Dhillon, and G. Caire, "Massive-MIMO meets HetNet: interference coordination through spatial blanking," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, May 2015.
- [19] T. Marzetta, "How much training is required for multiuser MIMO?," In *Proc. Asilomar Conf. Signals, Syst., and Comput. (ASILOMAR'06)*, Pacific Grove, CA, pp. 359–363, 2006.
- [20] Q. Sun, J. Wang, S. Jin, C. Xu, X. Gao and K. K. Wong, "Rate analysis and pilot reuse design for dense small cell networks," In *Proc. IEEE Int. Conf. Commun. (ICC'15)*, pp. 2845–2850, 2015.
- [21] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Englewood Cliffs, NJ: Prentice Hall, 1993.
- [22] H. Ngo, H. Suraweera, M. Matthaiou and E. Larsson, "Multipair massive MIMO full-duplex relaying with MRC/MRT processing," In *Proc. IEEE Int. Conf. Commun. (ICC'14)*, pp. 4818–4824, 2014.
- [23] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [24] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [25] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," <http://stanford.edu/boyd/cvx>, Feb. 2008.
- [26] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA), Further Advancements for E-UTRA Physical Layer Aspects," Mar. 2010 [Online]. Available: <ftp://ftp.3gpp.org>, TR 36.814 v9.0.0.
- [27] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.
- [28] H. Cramér, *Random Variables and Probability Distributions*, Cambridge, UK: Cambridge University Press, 1970.
- [29] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, pp. 1–182, June 2004.
- [30] N. H. Timm, "Applied multivariate analysis," *Springer*, 2002.
- [31] P. Graczyk, G. Letac, and H. Massam, "The complex Wishart distribution and the symmetric group," *Annals of Statistics*, vol. 31, pp. 287–309, 2003.